

EKONOMICKÁ UNIVERZITA V BRATISLAVE

FAKULTA HOSPODÁRSKEJ INFORMATIKY

Evidenčné číslo: 103006/I/2019/421000161824

**VYUŽITIE MARKOVOVÝCH REŤAZCOV V OBLASTI
AKTUÁRSTVA S PODPOROU JAZYKA R**

Diplomová práca

EKONOMICKÁ UNIVERZITA V BRATISLAVE
FAKULTA HOSPODÁRSKEJ INFORMATIKY

VYUŽITIE MARKOVÝCH REŤAZCOV V OBLASTI
AKTUÁRSTVA S PODPOROU JAZYKA R

Diplomová práca

Študijný program: Aktuársto

Študijný odbor: 6258 Kvantitatívne metódy v ekonómii

Školiace pracovisko: Katedra matematiky a aktuárstva

Vedúci záverečnej práce: doc. Mgr. Vladimír Mucha, PhD.



Ekonomická univerzita v Bratislave
Fakulta hospodárskej informatiky

ZADANIE ZÁVEREČNEJ PRÁCE

Meno a priezvisko študenta: Bc. Michaela Tkáčiková
Študijný program: aktuárstvo (Jednoodborové štúdium, inžiniersky II. st., denná forma)
Študijný odbor: 3.3.24 Kvantitatívne metódy v ekonómii
Typ záverečnej práce: Inžinierska záverečná práca
Jazyk záverečnej práce: slovenský
Sekundárny jazyk: anglický

Názov: Využitie Markovových reťazcov v oblasti aktuárstva s podporou jazyka R.

Anotácia: Teória Markovových reťazcov je založená na princípe náhodného procesu s časovým parametrom. Ten pre jednoduchosť možno charakterizovať ako postupnosť náhodných premenných, ktorých hodnoty nazývame stavmi. Pre Markovové reťazce platí, že pravdepodobnosť prechodu do nasledujúceho stavu závisí iba na súčasnom stave a nie na predchádzajúcich stavoch. Vzhľadom na fakt, že majú tieto reťazce uvedenú Markovovú vlastnosť, tak ich označujeme ako „bez pamäti“. Jednou z možností aplikovať túto teóriu v oblasti aktuárstva predstavuje Bonus – Malus systém, ktorý uplatňujú poisťovne v rámci PZP (povinného zmluvného poistenia zodpovednosti za škodu spôsobenú vedením motorového vozidla). Stavy, v ktorých sa poistenec nachádza nazývame bonus, resp. malus triedy. Výška poistného je závislá od daného stavu. Pre poisťovňu je dôležité vedieť, koľko poistencov bude po určitom čase v danej bonus - malus triede. Ďalšou oblasťou, ktorá sa opiera o modelovanie pomocou Markovových reťazcov je zdravotné a nemocenské poistenie. Poisťovňu bude v tomto prípade zaujímať počet poistencov, ktorí budú v danom stave (napríklad stav poistenca: mŕtvy, zdravý, chorý). Open source prostredie jazyka R prostredníctvom dostupných funkcií, resp. balíčkov (packages) predstavuje silný riešiteľský nástroj na realizáciu Markovovej analýzy, prípadne na generovania Markovových reťazcov v uvedených oblastiach z aktuárskej praxe.

Vedúci: doc. Mgr. Vladimír Mucha, PhD.
Katedra: KMA FHI - Katedra matematiky a aktuárstva FHI
Dátum zadania: 24.10.2017

Dátum schválenia: 03.11.2017

doc. RNDr. Jozef Fecenko, CSc.
vedúci katedry

Čestné vyhlásenie

Čestne vyhlasujem, že záverečnú prácu som vypracovala samostatne a uviedla všetku použitú literatúru.

Dátum: 30.4.2019

.....

ABSTRAKT

TKÁČIKOVÁ, Michaela: *Využitie Markovových reťazcov v oblasti aktuárstva s podporou jazyka R*. – Ekonomická univerzita v Bratislave. Fakulta hospodárskej informatiky, Katedra matematiky a aktuárstva. – Vedúci práce: doc. Mgr. Vladimír Mucha, PhD. – Bratislava: FHI, 2019, (57 s.)

Cieľom práce je zoznámiť sa s teóriou Markovových reťazcov a ukázať ich aplikáciu v oblasti aktuárstva. Práca obsahuje úvod do teórie Markovových reťazcov s diskretnou množinou stavov a diskretným a aj spojitým časom. Ďalej sme ukázali využitie software R ako hlavného riešiteľského nástroja pre našu problematiku. Teoretické poznatky sme aplikovali v oblasti životného aj neživotného poistenia. V oblasti neživotného poistenia to bola oblasť povinného zmluvného poistenia. Kde sme ukázali vytváranie matíc pravdepodobností prechodu pomocou faktora škodovosti a prínos vektora absolútnych pravdepodobností pre poisťovňu. V oblasti životného poistenia sme ukázali výpočet poisťného na základe nasimulovaných dát z matice pravdepodobností prechodu.

Kľúčové slová: Markovove reťazce, aktuárstvo, bonus malus systém, životné a nemocenské poistenie, matica pravdepodobností prechodu, vektor absolútnych pravdepodobností, software R

ABSTRACT

TKÁČIKOVÁ, Michaela: *The use of Markov chains in the field of actuarialism with the support of the language R*. - University of Economics in Bratislava. Faculty of Economic Informatics, Department of Mathematics and Actuarial Science.– Supervisor: doc. Mgr. Vladimír Mucha, PhD. – Bratislava: FHI, 2019, (57 s.)

The aim of the thesis is to get acquainted with the theory of Markov chains and to show their application in the field of actuarialism. The thesis contains an introduction to the theory of Markov chains with discrete set of states and discrete and continuous time. Furthermore, we have shown the use of R software as the main tool for solving our problems. We applied theoretical knowledge in both life and non-life insurance. In the area of non-life insurance, it was the area of compulsory contractual insurance. Where we showed the creation of transition probability matrices using the loss factor and the benefit of the absolute probability vector for the insurance company. In the area of life insurance, we have calculated the premium calculation based on simulated data from the transition probability matrix.

Keywords: Markov chains, actuariality, bonus malus system, life and sickness insurance, transition probability matrix, vector of absolute probabilities, software R

Obsah

Úvod	8
1. Súčasný stav riešenej problematiky doma a v zahraničí	10
1.1. Náhodný proces	10
<i>1.1.1. Základné vlastnosti</i>	<i>10</i>
<i>1.1.2. Distribučná funkcia náhodného procesu</i>	<i>11</i>
<i>1.1.3. Klasifikácia náhodného procesu</i>	<i>12</i>
1.1. Markovové reťazce	13
<i>1.1.1. Markovové reťazce s diskretným časom</i>	<i>13</i>
<i>1.1.2. Markovové reťazce so spojitým časom</i>	<i>18</i>
1.2. Modelovanie matice pravdepodobnosti prechodu	23
1.3. Bonus malus systém	25
1.4. Zdravotné a nemocenské poistenie	29
2. Cieľ práce	32
3. Metodika práce a metódy skúmania	33
3.1. Vytváranie objektov typu markovchain	34
3.2. Vykresľovanie objektov typu markovchain	35
3.3. Pravdepodobnostné operácie s objektami typu markovchain	36
3.4. Štatistická analýza a simulácie	37
4. Výsledky práce a diskusia	40
4.1. Aplikácia poznatkov na bonus malus systém	40
4.2. Aplikácia v oblasti životného a nemocenského poistenia	44
4.3. Diskusia	51
Záver	53
Zoznam použitej literatúry	54
Prílohy	56

Úvod

Teória Markovových reťazcov je založená na princípe náhodného procesu s časovým parametrom. Tento náhodný proces môžeme charakterizovať ako postupnosť náhodných premenných. Hodnoty, ktoré nadobúda náhodná premenná môžeme nazvať stavmi alebo triedami. Markovové reťazce musia spĺňať Markovovú vlastnosť, ktorá hovorí o tom, že pravdepodobnosť prechodu do nasledujúceho stavu závisí iba od súčasného stavu a nie od predchádzajúcich stavov. Nazývame ich aj Markovovými reťazcami „bez pamäti“. Časový parameter môže nadobúdať celočíselné hodnoty alebo hodnoty z intervalu. V prípade celočíselných hodnôt časového parametra hovoríme o Markovových reťazcoch s diskretným časom a v prípade, ak časový parameter bude nadobúdať hodnoty z intervalu hovoríme o Markovovom reťazci so spojitým časom. Ďalšie delenie reťazcov je podľa toho, či sa pravdepodobnosti prechodu s časom menia. Ak sa pravdepodobnosti prechodu s časom nemenia, hovoríme o homogénnom Markovovom reťazci a v opačnom prípade o nehomogénnom Markovovom reťazci. V tejto práci sa budeme venovať iba homogénnym Markovovým reťazcom, ale iba s diskretným časom.

Cieľom práce bude aplikovať Markovové reťazce na problematiku v aktuárstve. Markovove reťazce budeme aplikovať v oblasti životného aj neživotného poistenia.

Jedna z možností aplikácie Markovových reťazcov oblasti neživotného poistenia je bonus malus systém. Tento systém sa používa pri povinnom zmluvnom poistení zodpovednosti za škodu spôsobenú vedením motorového vozidla (PZP). Bonus malusová trieda (stav), v ktorých sa poistenec nachádza, určuje výšku jeho poistného. Ak poistenec spôsobí poistnú udalosť klesne do malusovej triedy a jeho poistné sa zvýši. Za bezškodový priebeh je poistenec odmenený zľavou z poistného. Pre poisťovňu je dôležité vedieť, koľko poistencov bude v bonus malusových triedach, aby vedela určiť výšku priemerného poistného a zostala solventná. Výpočtom stacionárneho vektora a priemerného poistného sa budeme venovať v praktickej časti tejto práce.

V oblasti životného poistenia budeme aplikovať Markovové reťazce na zdravotné a nemocenské poistenie. V tomto prípade budú poisťovňa zaujímať pravdepodobnosti prechodu medzi stavmi poistenca. Dáta, s ktorými budeme pracovať sú reálne dáta

z Talianska. Hovoria o mužskej populácii a o dlhodobó chorých. V praktickej časti sa zameriame na výpočet poistného pri nehomogénnych Markovových reťazcoch.

Silným riešiteľským nástrojom na realizáciu Markovových reťazcov je „open source“ prostredie jazyka R. V kapitole metodika práce a metódy skúmania si predstavíme balíčky a funkcie, ktoré nám pomôžu pri riešení a analyzovaní našej problematiky. V praktickej časti budeme aplikovať tieto riešiteľské nástroje na riešenu problematiku Markovových reťazcov aplikovaných na bonus malus systém a zdravotné a nemocenské poistenie.

1. Súčasný stav riešenej problematiky doma a v zahraničí

Využitie Markovových reťazcov v oblasti aktuárstva je veľmi široké. Tým, že sú flexibilné a jednoducho zrozumiteľné vidíme ich využitie v životom ako aj neživotnom poistení. Problematikou Markovových reťazcov sa zaoberá mnoho autorov. Celý matematický aparát je predstavený v knihe (Ethier 1986). Kniha sa zaoberá celkovou problematikou Markovových reťazcov. Sú v nej vysvetlené základné matematické aparáty, ktoré sú vysvetlené aj v tejto práci. V oblasti životného poistenia a nemocenského poistenia sa zaoberá (Haberman 1999) a (Dickson 2009). Kde sa opisujú základne princípy pri životnom poistení. V oblasti Markovových reťazcov a implementácie do softvéru R v súčasnosti robí najväčší výskum (Spedicato 2018). Vo svojej príručke opisuje balíčky a funkcie na ľahkú manipuláciu s Markovovými reťazcami.

V tejto kapitole si predstavíme a zadefinujeme Markovové reťazce a priblížime oblasti v poisťovníctve, kde sa využívajú. Priblížime si bonus malus systém a ako sa v ňom aplikujú Markovové reťazce. Ďalšou oblasťou, v ktorej si priblížime Markovové reťazce bude životné a nemocenské poistenie so zameraním na dlhodobo chorých a nesebestačných osôb.

1.1. Náhodný proces

V tejto časti uvedieme základné pojmy a vlastnosti, ktoré nám ozrejmi problematiku a priblížia ich používanie. Markovové reťazce, ktoré sú hlavnou náplňou práce, tvoria rozsiahlu triedu náhodných procesov, preto je im venovaná úvodná kapitola. Základy stochastického procesu môžeme nájsť v skriptách (Havrda 1981) a (Prášková 2005). Úvod do teórie pravdepodobnosti vrátane Markovových reťazcov je v (Grinstead 1997) a podrobnejší výklad teórie pravdepodobnosti a stochastických procesov je v (Knill 2011).

1.1.1. Základné vlastnosti

Náhodný proces môžeme chápať ako zovšeobecnenie náhodnej veličiny. Často je náhodný proces postupnosťou náhodných veličín v čase a zatiaľ čo náhodná veličina je

reálna funkcia elementárneho javu, náhodný proces je navyše aj funkciou času. Náhodné procesy tak vyjadrujú dynamiku náhodných javov.

Príkladom náhodného procesu môže byť Brownov pohyb hmotnej častice, postupnosť náhodných hodov hracou kockou, zmeny v počte zákazníkov čakajúcich na obsluhu a podobne.

Namiesto pojmu náhodný proces taktiež používame pojem stochastický proces alebo pravdepodobnostný proces.

Definícia 1.1. Nech $(\Omega, \mathcal{S}, \mathcal{P})$ je pravdepodobnostný priestor a nech T je neprázdnu podmnožinou množín reálnych čísel $\mathbb{R}$. Potom súbor náhodných veličín $\{X_t, t \in T\}$ definovaných na $(\Omega, \mathcal{S}, \mathcal{P})$ sa nazýva reálny náhodný proces (Grinstead 1997)

Ako už bolo povedané na náhodný proces sa môžeme pozerat' ako na funkciu dvoch premenných – elementárneho javu ω a času t . Pre pevne zvolené $t \in T$ je $X_t(\cdot)$ náhodná veličina definovaná na Ω . Pre pevne zvolené $\omega \in \Omega$ je $X_{(\cdot)}(\omega)$ je reálnou funkciou času $t \in T$. Funkcia $X_{(\cdot)}(\omega)$ sa nazýva trajektóriou alebo taktiež realizáciou náhodného procesu.

1.1.2. Distribučná funkcia náhodného procesu

Ďalej definujeme distribučnú funkciu náhodného procesu.

Definícia 1.2. Nech $\{X_t, t \in T\}$ je náhodný proces, $n \in \mathbb{N}$, $t_1, t_2, \dots, t_n \in T$ a $(X_{t_1}, X_{t_2}, \dots, X_{t_n})$ náhodný vektor, ktorého distribučná funkcia je predpisom $F_{t_1, t_2, \dots, t_n}(x_1, x_2, \dots, x_n) = P(X_{t_1} < x_1, X_{t_2} < x_2, \dots, X_{t_n} < x_n)$.

Ak sú pre systém distribučných funkcií $\{F_{t_1, t_2, \dots, t_n}(x_1, x_2, \dots, x_n)\}$ splnené nasledujúce tzv. Kolmogorové podmienky konzistencie

- pre ľubovoľnú permutáciu π množiny $\{1, 2, \dots, n\}$ platí

$$F_{t_{\pi(1)}, t_{\pi(2)}, \dots, t_{\pi(n)}}(x_{\pi(1)}, x_{\pi(2)}, \dots, x_{\pi(n)}) = F_{t_1, t_2, \dots, t_n}(x_1, x_2, \dots, x_n),$$

- n -rozmerná distribučná funkcia náhodného vektoru je marginálna distribučná funkcia $(n + 1)$ -rozmerného náhodného vektoru, tj.

$$\lim_{x_{n+1} \rightarrow \infty} F_{t_1, t_2, \dots, t_n, t_{n+1}}(x_1, x_2, \dots, x_n, x_{n+1}) = F_{t_1, t_2, \dots, t_n}(x_1, x_2, \dots, x_n).$$

Potom $F_{t_1, t_2, \dots, t_n}(x_1, x_2, \dots, x_n)$ sa nazýva n -rozmerná distribučná funkcia náhodného procesu $\{X_t, t \in T\}$.

Rozdelenie náhodného procesu je jednoznačne určené rozdelením všetkých konečne rozmerných náhodných vektorov $(X_{t_1}, X_{t_2}, \dots, X_{t_n})$. Každý náhodný proces určuje konzistentný systém distribučných funkcií (Grinstead 1997).

Veta 1.3 (Kolmogorov). Nech $\{F_{t_1, t_2, \dots, t_n}(x_1, x_2, \dots, x_n)\}$ je konzistentný systém distribučných funkcií. Potom existuje náhodný proces $\{X_t, t \in T\}$ taký, že pre každé $n \in \mathbb{N}$ ľubovoľné $t_1, t_2, \dots, t_n \in T$ a ľubovoľné $x_1, x_2, \dots, x_n$ platí

$$P(X_{t_1} < x_1, X_{t_2} < x_2, \dots, X_{t_n} < x_n) = F_{t_1, t_2, \dots, t_n}(x_1, x_2, \dots, x_n)$$

Dôkaz môžeme nájsť v (Prášková 2005).

1.1.3. Klasifikácia náhodného procesu

Klasifikácia náhodného procesu sa odvíja od štruktúry množiny stavov S , vlastností množiny T a vzťahu závislosti medzi náhodnou veličinou X_t pre rôzne $t \in T$.

Podľa typu množiny T rozlišujeme náhodný proces

- s diskretným časom, ak je množina stavov T diskretná, to je ak $T = \mathbb{N}_0$ resp. $T = \mathbb{Z}$,
- so spojitým časom, ak je T interval, to je ak $T = [a, b] \subset \mathbb{R}$.

Množina stavov náhodného procesu S je množina hodnôt náhodných veličín X_t .

Podľa štruktúry množiny S rozlišujeme

- náhodný proces s diskretnými stavmi, ak množina hodnôt náhodného procesu je konečná alebo spočítateľná, vtedy všetky náhodné veličiny X_t nadobúdajú iba diskkrétne hodnoty
- náhodný proces so spojitými stavmi, ak náhodné veličiny X_t nadobúdajú hodnoty z intervalu

Hádzanie kockou je príkladom náhodného procesu s diskretným časom, ktorý je definovaný poradovými číslami jednotlivých hodov s diskretnou množinou stavov, ktorá je tvorená počtom bodiek na kocke a nadobúda hodnoty $S = \{1, 2, 3, 4, 5, 6\}$. Ak sledujeme vývoj

počtu zákazníkov v rade, jedná sa o náhodný proces so spojitým časom, kde množina T je daná sledovaným obdobím, a množina stavov je $S = \{0, 1, 2, \dots\}$. Pokiaľ budeme zaznamenávať vonkajšiu teplotu v pravidelných intervaloch, ide o príklad náhodného procesu s diskretným časom, ktorým je poradové číslo nameraných hodnôt a spojitou množinou stavov (vonkajšia teplota). Vo všeobecnosti proces vývoja vonkajšej teploty môžeme chápať ako náhodný proces so spojitým časom a spojitou množinou stavov.

Ďalej uvedieme základné príklady náhodných procesov podľa typu závislosti medzi náhodnými veličinami pri rôznych časoch (Havrda 1981):

- i. Procesy s nezávislými prírastkami. Nazývame taký proces pre ktorý platí, že náhodné veličiny $X_{t_2} - X_{t_1}, X_{t_3} - X_{t_2}, \dots, X_{t_n} - X_{t_{n-1}}$ sú nezávislé pre každé $t_j \in T$, $j = 1, 2, \dots, n$ a $t_1 < t_2 < \dots < t_{n-1} < t_n$. Ak napríklad $Y_1, Y_2, \dots, Y_n$ sú hodnoty náhodného výberu, položíme $X_{t_1} = Y_1, X_{t_2} = Y_1 + Y_2, \dots, X_{t_n} = Y_1 + Y_2 + \dots + Y_n$. Potom ide o proces s nezávislými prírastkami, kde $T = \{1, 2, \dots, n\}$ a $X_{t_n} - X_{t_{n-1}} = Y_{n+1}, n = 1, 2, \dots$.
- ii. Markovové procesy. Náhodný proces $\{X_t, t \in T\}$ nazveme Markovovým procesom, ak má takú vlastnosť, že jeho stav v budúcom čase $t > s$ závisí iba od stavu v súčasnom čase s a nie na stavoch v predošlom čase $u < s$, kde $s, t, u \in T$, je to teda proces „bez pamäti“. Formálne

$$P(X_t < a | X_{t_n} = i_n, \dots, X_{t_1} = i_1, X_{t_0} = i_0) = P(X_t < a | X_{t_n} = i_n)$$

pre každé $t_0 < t_1 < \dots < t_n < t, a \in \mathbb{R}$, pre ktoré platí

$$P(X_{t_n} = i_n, \dots, X_{t_1} = i_1, X_{t_0} = i_0) > 0.$$

1.2. Markovové reťazce

Markovové procesy sú pomenované po Andrejovi Andrejevičovi Markovovi, ruskom matematikovi, ktorý sa zaoberal náhodnými procesmi.

Hlavným zdrojom o Markovových reťazcoch pre túto kapitolu boli zdroje (Havrda 1981), (Křivý 2005) a (Prášková 2005). Stochastický a Markovov proces môžeme nájsť v knihe (Ethier 1986).

Bez obmedzenia na všeobecnosti budeme ďalej pojem Markovov reťazec používať pre Markovov proces s diskretným časom. V takom prípade náhodné veličiny $\{X_t, t \in T\}$ nadobúdajú iba celočíselné hodnoty, ktoré tvoria množinu stavov S Markovovho reťazca. S je konečné alebo spočítateľné, teda $S = \{1, 2, \dots, N\}$ alebo $S = \{1, 2, \dots\}$.

1.2.1. Markovové reťazce s diskretným časom

Definícia 1.3. Nech $\{X_n, n \in T\}$ je náhodný proces s diskretnými stavmi a $S \in \mathbb{N}_0$ je množina jeho stavov. Hovorme, že X je *Markovov reťazec s diskretným časom* a množinou S , ak je $T = \mathbb{N}_0$ a je splnená podmienka

$$P(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) = P(X_{n+1} = j | X_n = i) \quad (2.1)$$

Pre všetky $n \in T$ a všetky stavy $i_0, \dots, i_{n-1}, i, j \in S$, také že

$$P(X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) > 0.$$

Podmienkou (2.1) vyjadrujeme takzvanú Markovovú vlastnosť, ktorá znamená, že pravdepodobnosť s akou proces prejde do nejakého stavu v nasledujúcom čase $n + 1$ závisí iba na jeho stave v súčasnom čase n a nezáleží na tom, v akom stave sa náhodný proces nachádzal predtým, tj. v časoch $n - 1, n - 2, \dots, 1, 0$ (Ethier 1986).

Zmena stavu procesu sa nazýva prechodom. Reťazec sa zo stavu i v nasledujúcom časovom okamihu dostane do stavu j s pravdepodobnosťou $p_{ij}(n, n + 1)$. Táto pravdepodobnosť sa nazýva pravdepodobnosť prechodu 1. rádu zo stavu i do stavu j , a teda

$$p_{ij}(n, n + 1) = P(X_{n+1} = j | X_n = i)$$

Všeobecná podmienka pravdepodobnosti

$$p_{ij}(n, n + m) = P(X_{n+m} = j | X_n = i)$$

sa nazýva pravdepodobnosť prechodu m -tého radu, tj. pravdepodobnosť prechodu zo stavu i v čase n do stavu j v čase $n + m$.

Definícia 1.4. Nech $\{X_n, n \in \mathbb{N}_0\}$ je Markovov reťazec s diskretným časom a množinou stavov S . Hovorme, že X je *homogénny Markovov reťazec*, ak podmienené

pravdepodobnosti $p_{ij}(n, n + m)$ nezávisia na východiskovom čase n , ale len na rozdiel stavov m . Potom značíme $p_{ij}(n, n + 1)$, resp. $p_{ij}^{(m)}(n, n + m)$. V opačnom prípade hovoríme, že Markovov reťazec je nehomogénny (Ethier 1986).

V ďalšej časti práce budeme predpokladať len homogénny Markovov reťazec s diskrétnym časom.

Podmienené pravdepodobnosti p_{ij} môžeme usporiadať do štvorcovej matice. $\mathbf{P} = \{p_{ij}, i, j \in S\}$, ktorú nazývame *matica pravdepodobností prechodu*. Pre každé $i, j \in S$ platí

$$p_{ij} \geq 0, \sum_{j \in S} p_{ij} = 1$$

ide o *stochastickú maticu* (prvky matice sú nezáporné a súčet prvkov v každom riadku matice je rovný jednej).

Veta 2.3 (pravdepodobnosť prechodu n -tého rádu) Nech $\{X_n, n \in \mathbb{N}_0\}$ je homogénny Markovov reťazec s maticou pravdepodobností prechodu $\mathbf{P}$. Nech $p_{ii}^{(0)} = 1, p_{ij}^{(0)} = 0$ pre $i \neq j$ a $p_{ij}^{(1)} = p_{ij}$. Potom pre podmienené pravdepodobnosti n -tého rádu $p_{ii}^{(n)} = P(X_{n+m} = j | X_n = i)$ platí

$$p_{ij}^{(n)} = \sum_{k \in S} p_{ik}^{(n-1)} p_{kj}, j, k \in S \quad (2.2)$$

Pre všetky $n, m \in \mathbb{N}_0$ a $P(X_m = i) > 0$.

Dôkaz je uvedený v (Prášková 2005).

Zo vzťahu (2.2) vyplýva, že matica $\mathbf{P}^{(n)}$ prvkov $p_{ij}^{(n)}$ je stochastická matica a platí

$$\mathbf{P}^{(n)} = \mathbf{P}^{(n-1)} \cdot \mathbf{P} = \mathbf{P} \cdot \mathbf{P}^{(n-1)} = \mathbf{P}^n.$$

Vzťah (2.2) môžeme zovšeobecniť

$$p_{ij}^{(m+n)} = \sum_{k \in S} p_{ik}^{(m)} p_{kj}^{(n)}$$

pre všetky $m, n \in \mathbb{N}_0$, ktorý sa nazýva Chapmanová – Kolmogorová rovnosť. Maticovo môžeme rovnosť vyjadriť ako $\mathbf{P}^{(m+n)} = \mathbf{P}^{(m)} \cdot \mathbf{P}^{(n)}$.

Pravdepodobnosti, s akými Markovov reťazec nadobúda jednotlivé stavy v počiatočnom čase $p_i = P(X_0 = i), i \in S$, určuje počiatočné rozdelenie Markovovho reťazca $\{X_n, n \in \mathbb{N}_0\}$. Toto rozdelenie budeme značiť $\mathbf{p} = \{p_i, i \in S\}$.

Platí

$$p_i \geq 0, \sum_{i \in S} p_i = 1.$$

Nepodmienené pravdepodobnosti $p_i(n) = P(X_n = i)$ nazývame absolútnou pravdepodobnosťou v čase n . Vektor absolútnych pravdepodobností budeme značiť $\mathbf{p}(n) = \{p_i(n), i \in S\}$.

Pre tieto nepodmienené pravdepodobnosti platí vzťah

$$\begin{aligned} p_j(n) &= P(X_n = j) = \sum_{k \in S} P(X_0 = k, X_n = j) \\ &= \sum_{k \in S} P(X_0 = k | X_n = j) P(X_0 = k) \\ &= \sum_{k \in S} p_k p_{kj}^n. \end{aligned}$$

Vektorovo môžeme vzťah medzi vektorom absolútnych pravdepodobností a počiatočným rozdelením vyjadriť nasledovne:

$$\mathbf{p}(n)^T = \mathbf{p}^T \mathbf{P}^n, n \in \mathbb{N}_0$$

Najskôr klasifikujeme stavy Markovovho reťazca podľa typu prechodu medzi stavmi.

Definícia 2.4. Hovorme že stav j je dosiahnuteľný zo stavu i , ak existuje $n \in \mathbb{N}_0$ také, že $p_{ij}^{(n)} > 0$, tento jav budeme značiť $i \rightarrow j$. Hovoríme, že stavy i a j sú vzájomne dosiahnuteľné, ak stav i je dosiahnuteľný z j a naopak, čo budeme značiť $i \leftrightarrow j$ (Ethier 1986).

Definícia 2.5. Hovorme, že neprázdna množina stavov $C \subseteq S$ je uzatvorená, ak pre každé $i \in C$ a každé j také, že $i \rightarrow j$, tak $j \in C$ (Ethier 1986).

Definícia 2.6. Stav i sa nazýva absorpčný, ak jednobodová množina $\{i\}$ je uzavretá, to je pravdepodobnosť $p_{ii} = 1$.

Definícia 2.7. Markovov reťazec je nerozložiteľný, ak každé dva jeho stavy sú vzájomne dosiahnuteľné. V opačnom prípade je rozložiteľný.

Definícia 2.8. Hovorme, že stav i je periodický s periódou $d > 1$, ak d je najväčší spoločný deliteľ množiny $\{n \in \mathbb{N}: p_{ii}^{(n)} > 0\}$. Ak najväčším spoločným deliteľom je $d = 1$, potom sa stav i nazýva aperiodický.

Dosiahnuteľnosť medzi jednotlivými stavmi sa dá znázorniť orientovaným grafom, v ktorom sú jednotlivé stavy reprezentované jeho vrcholmi a orientovaná hrana vedie medzi vrcholmi i a vrcholom j práve vtedy keď $i \neq j$ a $p_{ij} > 0$. Stav k je potom dosiahnuteľný zo stavu l práve vtedy, ak z vrcholu l vedie orientovaná cesta do vrcholu k .

Uvažujeme Markovov reťazec, ktorý sa v čase m nachádza v stave i a po n krokoch sa dostane práve do stavu j . Pravdepodobnosť, že sa tak stane, nazveme pravdepodobnosťou prvého dosiahnutia stavu j zo stavu i po n krokoch a označme ju $f_{ij}^{(n)}$, potom

$$f_{ij}^{(n)} = P(X_{m+n} = j, X_{m+n-1} \neq j, \dots, X_{m+1} \neq j | X_m = i)$$

Pravdepodobnosť $f_{ii}^{(n)}$ nazveme pravdepodobnosťou prvého návratu do stavu i po n krokoch.

Pravdepodobnosť f_{ij} , s akou reťazec, ktorý sa nachádza v čase m v stave i , vstúpi po konečne veľa krokoch po prvý krát do stavu j nazveme pravdepodobnosťou prvého dosiahnuteľného stavu j zo stavu i (po konečne veľa krokoch), a platí pre ne vzťah

$$f_{ij} = \sum_{n=1}^{\infty} f_{ij}^{(n)}, i, j \in S$$

Podobne pravdepodobnosť f_{ii} nazývame pravdepodobnosťou prvého návratu do stavu i (po konečne veľa krokoch).

Definícia 2.9. Nech $\{X_n, n \in \mathbb{N}\}$ je Markovov reťazec s konečnou alebo spočítateľnou množinou stavov S . Náhodnú veličinu $\tau_j(1) = \inf\{n > 0: X_n = j, i, j \in S\}$ nazveme časom (dobou) prvého návratu do stavu j .

Veličinou $\tau_j(1)$ značíme počet krokov po ktorých sa reťazec po prvý krát vráti do stavu j , nadobúda teda hodnoty $1, 2, \dots$, v prípade, že také n , v ktorom by sa reťazec vrátil do stavu j , neexistuje, kladieme $\tau_j(1) = \infty$. Náhodná veličina $\tau_j(k)$, k -ty návrat do stavu j , sa rovná času, v ktorom nastane k -ty návrat do stavu j alebo ∞ v prípade, že sa Markovov reťazec do stavu j nevráti aspoň k -krát.

Náhodná veličina $\tau_j(1)$ má pravdepodobnostné rozdelenie $\{f_j^{(n)}\}$, tj.

$$f_j^{(n)} = P_i(\tau_j(1) = n).$$

Definícia 2.10. Nech $\{X_n\}_{n \geq 1}$ je Markovov reťazec s konečnou alebo spočítateľnou množinou stavov S . Stav $i \in S$ nazveme

- i. trvalým, ak pravdepodobnosť prvého návratu do stavu j je 1, $P(\tau_j(1) < \infty) = 1$,
- ii. trvalým nenulovým, ak je stav trvalý a zároveň $E_j \tau_j(1) < \infty$,
- iii. trvalým nulovým, ak je stav trvalý a zároveň $E_j \tau_j(1) = \infty$,
- iv. prechodným, ak pravdepodobnosť, že sa reťazec do stavu j nikdy nevráti nie je nenulová, resp. návrat do stavu j není istý jav, teda $P(\tau_j(1) < \infty) < 1$.

1.2.2. Markovové reťazce so spojitým časom

Ako už bolo povedané vyššie o náhodný proces so spojitým časom sa jedná, ak množina T je interval a budeme uvažovať $T = [0, \infty)$.

Definícia 2.11. Nech $\{X_t, t \in T\}$ je náhodný proces s diskretnými stavmi a množinou týchto stavov S . Hovoríme, že X_t je Markovov reťazec so spojitým časom a spočítateľnou množinou stavov S , ak $T = [0, \infty)$ a je splnená podmienka

$$P(X_t = j | X_s = i, X_{t_n} = i_n, \dots, X_{t_1} = i_1) = P(X_t = j | X_s = i) \quad (2.3)$$

pre všetky $0 < t_1 < \dots < t_n < s < t$ a všetky stavy $i_1, \dots, i_n, i, j \in S$ platí, že $P(X_{t_1} = i_1, \dots, X_{t_n} = i_n, X_s = i, X_t = j) > 0$. Podmienka (2.3) vyjadruje opäť Markovovú vlastnosť náhodného procesu (Ethier 1986).

Podobne ako pri Markovových reťazcoch s diskretným časom bude podmienenou pravdepodobnosťou $P(X_t = j | X_s = i)$, že reťazec, ktorý je v čase s v stave i , bude v čase $t, t > s$, v stave j , značiť $p_{ij}(s, t)$ a budeme ju nazývať pravdepodobnosťou prechodu zo stavu i do stavu j v čase t . Zjavne platí

$$p_{ij}(s, t) \geq 0, \sum_{j \in S} p_{ij}(s, t) = 1$$

Taktiež absolútne pravdepodobnosti $p_j(t)$ sú pravdepodobnosti rozdelenia náhodnej veličiny X_t , potom $p_i(t) = P(X_t = i), i \in S, t \in T$ a počiatočné pravdepodobnosti p_j sú pravdepodobnosti $p_j = p_j(0) = P(X_0 = j), j \in S$.

Pravdepodobnosť $p_{ij}(s, t)$ sa dá pri pevných s, t usporiadať do stochastickej matice pravdepodobností prechodu $\mathbf{P}(s, t)$, definujeme $p_{ii}(0) = 1, p_{ij}(0) = 0$ pre $i \neq j$ a teda $\mathbf{P}(0) = \mathbb{I}$.

Veta 2.12. Nech $\{X_t, t \geq 0\}$ je Markovov reťazec so spojitým časom. Pre absolútne pravdepodobnosti $p_i(t)$ a pravdepodobnosti prechodu $p_{ij}(s, t)$ platí

$$p_j(t) = \sum_{k \in S} p_k(s) p_{kj}(s, t), s < t, j \in S \quad (2.4)$$

$$p_{ij}(s, t) = \sum_{k \in S} p_{ik}(s, l) p_{kj}(l, t), s < l < t, i, j \in S \quad (2.5)$$

Dôkaz podľa (Havrda 1981).

Definícia 2.13. Nech $\{X_t, t \geq 0\}$ je Markovov reťazec so spojitým časom a množinou stavov S . Hovorme, že reťazec je homogénny, ak platí, že pravdepodobnosti prechodu $p_{ij}(s, v)$ závisí len na rozdieli $v - s = t$, tj. $p_{ij}(s, v) = p_{ij}(t)$, pre všetky $v > s \geq 0$ a pre každé $i, j \in S$. V opačnom prípade hovoríme, že Markovov reťazec je nehomogénny.

Vzťahy (2.4), (2.5) môžeme pre homogénny Markovov reťazec definovať ako

$$p_j(s + t) = \sum_{k \in S} p_k(s) p_{kj}(t), s, t \geq 0, j \in S \quad (2.6)$$

$$p_{ij}(s + t) = \sum_{k \in S} p_{ik}(s) p_{kj}(t), s, t \geq 0, i, j \in S \quad (2.7)$$

Matica

$$\mathbf{P}(s + t) = \mathbf{P}(s)\mathbf{P}(t)$$

Jedná sa opäť o Chapmanovu – Kolmogorovovu rovnicu.

V prípade homogénnych Markovových reťazcov a diskrétnym časom bola matica pravdepodobností prechodu závislá iba na počte „časových intervalov“ nie však na ich dĺžke. Na druhej strane, matica pravdepodobností prechodu Markovovho reťazca so spojitým časom je práve funkciou dĺžky časových intervalov medzi zmenami stavov. Pokiaľ je Markovov reťazec v stave i v čase s a do ďalšieho stavu sa dostane v čase $s + t$, je t doba zotrvania v stave i .

Doba zotrvania v stave i je vzhľadom na markovové vlastnosti nezávislá na dobe zotrvania v predchádzajúcich stavoch, teda je „bez pamäti“, a teda je to náhodná veličina s exponenciálnym rozdelením s parametrom μ_i .

Pozrime sa ďalej na vzťah medzi μ_i a pravdepodobnosťami prechodu $p_{ij}(t)$.

Veta 2.14. Nech $\{X_t, t \geq 0\}$ je Markovov reťazec so spojitým časom a množinou stavov S a maticou pravdepodobností prechodu $\mathbf{P}(t)$ s prvkami $p_{ij}(t)$, pre ktoré platí

$$\lim_{t \rightarrow 0^+} p_{ij}(t) = \delta_{ij}$$

kde dodefinujeme $p_{ij}(0) = \delta_{ij}$, teda $p_{ij}(t)$ sú pre $t \in T$ spojitú sprava v bode nula pre všetky $i, j \in S$. Potom pre každé $i \in S$ existuje limita

(2.8)

$$\mu_i = \lim_{t \rightarrow 0^+} \frac{1 - p_{ij}(t)}{t}$$

ktorá však môže byť aj nekonečná. A pre každé $i, j \in S, i \neq j$ existujú konečné limity

$$\mu_{ij} = \lim_{t \rightarrow 0^+} \frac{p_{ij}(t)}{t} \quad (2.9)$$

Dôkaz podľa (Chung 1967).

Definícia 2.15. Hovorme, že homogénny Markovov reťazec je konzistentný, ak pre všetky $i \in S$ platí

$$\mu_i = \sum_{i \neq j} \mu_{ij} \quad (2.10)$$

Pre reťaze s konečnou množinou stavov platí vzťah (2.10) vždy, čo hovorí nasledujúca veta (Ethier 1986).

Veta 2.16. Homogénny Markovov reťazec s konečným počtom stavov je konzistentný.

Dôkaz podľa (Havrdá 1981).

Ak uvažujeme len konzistentný Markovov reťazec s konečným počtom stavov.

Definícia 2.17. Nezáporné čísla μ_{ij} definované v 2.14 sa nazývajú intenzity prechodu zo stavu i do stavu j , nezáporné číslo μ_i sa nazýva celková intenzita prechodu zo stavu i . Matica $Q = \{\mu_{ij}, i, j \in S\}$, kde $\mu_{ii} = -\mu_i$, sa nazýva matica intenzít prechodu alebo generujúca matica procesu.

Riadkové súčty matice intenzít pri konzistentnom Markovovom reťazci sa zrejme rovnajú nule.

Ako sme už uviedli v úvode odstavca je μ_i parameter exponenciálneho rozdelenia náhodnej veličiny doby zotrvania v stave i .

Veta 2.18. Ak je $\mu_i = 0$, potom $\mu_{ii}(t) = 1$ pre všetky $t \geq 0$. Ak je $\mu_i \in (0, \infty)$, potom doba zotrvania v stave i má exponenciálne rozdelenie so strednou hodnotou $\frac{1}{\mu_i}$.

Dôkaz podľa (Prášková 2005).

Definícia 2.19. Hovorme, že stav $i \in S$ je absorpčný, ak $\mu_i = 0$. Hovorme, že stav $i \in S$ je stabilný, ak $\mu_i \in (0, \infty)$ a nestabilný, ak $\mu_i = \infty$.

Pravdepodobnosť, že behom intervalu $(s, s + h)$, $h > 0$ nastane prechod zo stavu i do nejakého iného stavu je $\mu_i h + o(h)$, kde μ_i je parameter exponenciálneho rozdelenia doby zotrvania v stave i . Označme pravdepodobnosť, že reťazec prejde zo stavu i najprv do stavu j ako μ_{ij}^* . Pre podmienenú pravdepodobnosť prechodu do stavu j platí

$$P(X_{s+h} = j | X_s = i) = \mu_i \mu_{ij}^* h + o(h) = \mu_{ij} h + o(h)$$

z čoho vyplýva pre $\mu_i \neq 0$.

$$\mu_{ij}^* = \frac{\mu_{ij}}{\mu_i} \text{ pre } i \neq j \text{ a}$$

$$\mu_{ii}^* = 0.$$

Pokiaľ je $\mu_i = 0$ zotrvanie reťazca v stave i nekonečné dlhé ide o absorpčný stav, a preto pre pravdepodobnosti, že reťazec prejde do nejakého iného stavu μ_{ij}^* , platí

$$\mu_{ij}^* = 0 \text{ pre } i \neq j \text{ a}$$

$$\mu_{ii}^* = 1.$$

Prvky μ_{ij}^* môžeme usporiadať do matice $\mathbf{Q}^*$.

Označme časové okamžiky, v ktorých príde k zmene stavu reťazca $\{X_t, t \geq 0\}$ ako $s_0, s_1, \dots$ a definujme $Y_0 = X_0, Y_1 = X_{s_1}, \dots, Y_n = X_{s_n}, \dots$. Potom $\{Y_n, n \in \mathbb{N}_0\}$ je homogénny Markovov reťazec s diskretným časom s množinou stavov S a maticou pravdepodobností prechodu $\mathbf{Q}^*$. Reťazec $\{Y_n, n \in \mathbb{N}_0\}$ sa nazýva vnorený diskretný reťazec procesu $\{X_t, t \geq 0\}$.

Predpoklad známej matice intenzít prechodu $\mathbf{Q}$ môžeme v prípade konečnej množiny stavov vždy odvodiť maticu pravdepodobností prechodu $\mathbf{P}(t)$ Markovovho reťazca. V prípade spočítateľnej množiny stavov Markovovho reťazca je hľadanie jeho matice pravdepodobností prechodu $\mathbf{P}(t)$ zložitejšie a výsledok nemusí byť jednoznačný.

Veta 2.20 Pre homogénny Markovov reťazec X_t platí:

(1) Systém priamych (prospektívnych) Kolmogorovových rovníc

$$p'_{ij}(t) = \sum_{i \in S} p_{ik}(t) \mu_{ik}, i, j \in S, \quad (2.11)$$

maticovo $\mathbf{P}'(t) = \mathbf{P}(t)\mathbf{Q}$.

(2) Systém opačných (retrospektívnych) Kolmogorovových rovníc

$$p'_{ij}(t) = \sum_{i \in S} \mu_{ik} p_{ik}(t), i, j \in S, \quad (2.12)$$

maticovo $\mathbf{P}'(t) = \mathbf{Q}\mathbf{P}(t)$.

Dôkaz bodla (Prášková 2005).

Veta 2.21 Nech máme maticu $\mathbf{Q}$ s konečnou dimenziou, pre jej prvky platí $\mu_{ij} \geq 0$, $i \neq j$ a $\mu_{ii} = -\sum_{i \neq j} \mu_{ij}$. Potom existuje jediné riešenie sústavy priamych a opačných Kolmogorovových rovníc uvedených vo vete 2.20, rovnaké pre obe sústavy, ktorá vyhovuje počiatočnej podmienke $\mathbf{P}(0) = \mathbf{I}$. Toto riešenie predstavuje maticu pravdepodobností prechodu $\mathbf{P}(t)$ homogénneho Markovovho reťazca $\{X_t, t \geq 0\}$ s konečnou množinou stavov.

Dôkaz nájdete (Prášková 2005).

Sústavy Kolmogorovových rovníc uvedené vo vete 2.20 sú sústavou homogénnych lineárnych rovníc 1. radu s konštantnými koeficientami. Všeobecné riešenie má tvar

$$\mathbf{P}(t) = \mathbf{P}(0)e^{\mathbf{Q}t},$$

kde výraz $e^{\mathbf{Q}t}$ je skrátenejší zápis pre

$$\sum_{k=0}^{\infty} \frac{\mathbf{Q}^k t^k}{k!}$$

Ak riešenie vyhovuje počiatočnej podmienke $\mathbf{P}(0) = \mathbf{I}$, môžeme ho zapísať v tvare

$$\mathbf{P}(t) = e^{\mathbf{Q}t}. \quad (2.13)$$

V prípade nehomogénneho Markovovho reťazca pravdepodobnosti prechodu $p_{ij}(s, s+t)$ závisia nie len na veľkosti intervalu t , ale aj na počiatočnom časovom okamihu

s. Predpokladajme, že pre každé $i, j \in S$ v každom časovom okamžiku s existuje intenzita prechodu tak, ako je definované vo vete 2.8. Intenzity prechodu sa tak stanú funkciou času s a teda

$$\mu_{ij}(s) = \lim_{t \rightarrow 0^+} \frac{p_{ij}(s, s+t)}{t}, i \neq j, i, j \in S.$$

Pretože predpokladáme, že intenzita prechodu μ_{ij} existuje pre všetky dvojice i, j , môžeme definovať $\mu_{ii}(s) = -\sum_{i \neq j} \mu_{ij}(s)$. Prvky $\mu_{ij}(s)$ môžeme usporiadať do matice $\mathbf{Q}(s)$. V prípade konečnej množiny stavov S pre Kolmogorovové diferenciálne rovnice s počiatočnou podmienkou $\mathbf{P}(s, s) = \mathbf{I}$ platí

$$\frac{\partial \mathbf{P}(s, t)}{\partial} = \lim_{h \rightarrow 0} \frac{\mathbf{P}(s, t+h) - \mathbf{P}(s, t)}{h} = \mathbf{P}(s, t) \lim_{h \rightarrow 0} \frac{\mathbf{P}(t, t+h) - \mathbf{I}}{h} = \mathbf{P}(s, t) \mathbf{Q}(t),$$

teda majú tvar (derivácia značí parciálnu deriváciu podľa t)

$$\mathbf{P}'(s, t) = \mathbf{P}(s, t) \mathbf{Q}(t), \quad (2.14)$$

$$\mathbf{P}'(s, t) = -\mathbf{Q}(t) \mathbf{P}(s, t). \quad (2.15)$$

Ich riešenie nie je už také jednoduché ako pri homogénnych Markovových reťazcoch.

Definícia 2.22. Definujeme kumulatívne intenzity prechodu $a_{ij}(t)$ že stavy i v čase s do stavu j v čase $t, j, i \in S$ ako

$$a_{ij}(t) = \int_0^t \mu_{ij}(s) ds,$$

$$a_{ii} = -\sum_{j \neq i} \mu_{ij}(t).$$

Kumulatívne intenzity prechodu môžeme usporiadať do matice $\mathbf{A}(t)$ (Ethier 1986).

Riešením Kolmogorovovej rovnice 2.14 je potom

$$\mathbf{P}(s, t) = \prod_{(s, t >)} (\mathbf{I} + d\mathbf{A}(t))$$

$$= \lim_{\max(t_i - t_{i-1})} \prod_{(s, t >)} (\mathbf{I} + \mathbf{A}(t_i) - \mathbf{A}(t_{i-1})),$$

kde $s = t_0 < \dots < t_n = t$ a $\mathbf{I}$ je jednotková matica s dimenziou stanovenou priestorom S .

1.3. Modelovanie matice pravdepodobnosti prechodu

Ak chceme použiť Markovove reťazce na modelovanie reálnej situácie, tak je veľká pravdepodobnosť, že nepoznáme maticu pravdepodobností prechodu. Musíme preto použiť ich odhady a vedieť ich štatistické vlastnosti. V tejto časti sa budeme venovať riešeniu tohto problému pre diskretný homogénny Markovov reťazec. Hlavným zdrojom pre túto časť je (Janková 2014).

Nech $\{X_n, n \in \mathbb{N}_0\}$ je Markovov reťazec s konečnou množinou stavov $S = \{1, 2, \dots\}$, ktoré ma počiatkové rozdelenie α a neznámu maticu pravdepodobností prechodu $\mathbf{P}$. Túto maticu budeme chcieť odhadnúť na základe pozorovaní v časoch $0, 1, \dots, n$.

Predpokladajme že reťazec $\{X_n, n \in \mathbb{N}_0\}$ nadobúda v časoch $0, 1, \dots, n$ nasledujúce hodnoty

$$X_0 = i_0, X_1 = i_1, \dots, X_n = i_n,$$

kde $i_0, i_1, \dots, i_n \in S$. Potom z Markovovej vlastnosti vieme

$$P(X_0 = i_0, X_1 = i_1, \dots, X_n = i_n) = p_{i_0} p_{i_0 i_1} \dots p_{i_{n-1} i_n} = p_{i_0} \prod_{i,j \in S} p_{ij}^{n_{ij}},$$

kde n_{ij} je realizácia náhodnej premennej N_{ij} ktoré predstavuje počet pozorovaných prechodov zo stavu i do stavu j v časoch $0, 1, \dots, n-1$. Matica $\mathbf{N} = (N_{ij})_{i,j \in S}$ je matica početností pozorovaných prechodov v časoch $0, 1, \dots, n$.

Veta 3.1. Ak poznáme $X_0 = i_0$ a maticu $\mathbf{N}$, tak združené rozdelenie náhodného vektora $(X_0, \dots, X_n)$ nezávisí od matice prechodu $\mathbf{P}$ Markovovho reťazca $\{X_n, n \in \mathbb{N}_0\}$, presnejšie

$$P(X_0 = i_0, X_1 = i_1, \dots, X_n = i_n | X_0, \mathbf{N}) \quad (3.1)$$

nezávisí od prvkov matice $\mathbf{P}$.

Dôkaz podľa (Janková 2014).

Na odhad prvkov $\hat{p}_{ik}$ matice $\mathbf{P}$ použijeme metódu maximálnej vierohodnosti. Vierohodnosť pozorovania $(i_0, \dots, i_n)$ je funkcia neznámych parametrov

$$p_{11}, \dots, p_{i \ m-1}, \dots, p_{m1}, \dots, p_{m \ m-1}.$$

Pravdepodobnosti p_{im} sa dajú dopočítať, pretože riadky matice $\mathbf{P}$ tvoria rozdelenie pravdepodobnosti. Máme vierohodnú funkciu

$$\begin{aligned} L(\mathbf{P}) &= \alpha_{i_0} \prod_{i,j \in S} p_{ij}^{n_{ij}} \\ &= \alpha_{i_0} \prod_{i=1}^m \prod_{j=1}^{m-1} p_{ij}^{n_{ij}} \left(1 - \sum_{k=1}^{m-1} p_{ik}\right)^{n_{im}}, \end{aligned}$$

jej parametre spĺňajú ohraničenie

$$\sum_{j=1}^m p_{ij} = 1$$

pre každé i .

V nasledujúcom kroku dostaneme

$$\frac{\partial \ln L}{\partial p_{ij}} = \frac{n_{ij}}{p_{ij}} - \frac{n_{im}}{1 - \sum_{j=1}^{m-1} p_{ij}} = 0. \quad (3.2)$$

Ďalej označme

$$n_i = \sum_{j \in S} n_{ij}.$$

Ak $n_i > 0$ pre nejaké $i \in S$, tak ak riešime (3.2) dostaneme

$$p_{ik} = \frac{n_{ik}}{n_i} \quad (3.3)$$

pre $k \in S$. Odhad metódou maximálnej vierohodnosti pre p_{ik} , $i, k \in S$, je

$$\hat{p}_{ik} = \frac{N_{ik}}{N_i}, \quad (3.4)$$

kde N_{ik} je celkový počet pozorovaných prechodov z i do k a $N_i = \sum_{j \in S} N_{ij} \neq 0$. Ak $N_i = 0$ tak potom aj $\hat{p}_{ik} = 0$.

1.4. Bonus malus systém

Poisťovne na Slovensku používajú k stanoveniu výšky poistného v povinnom zmluvnom poistení za škodu spôsobenú vedením motorového vozidla systém bonus malus.

Ideou bonus malus systému je odmeňovať vodičov za bezškodový priebeh poistenia zľavou z poistného a naopak prirážkou k poistnému za škodovú udalosť.

Každý vlastník motorového vozidla, ktorý ho chce prevádzkovať na cestnej komunikácii na Slovensku, musí mať toto poistenie podpísané podľa zákona, no môže si vybrať v akej poisťovni sa poistiť.

Systém bonus malus prišiel do platnosti 1.1.2002 v zákone o povinnom zmluvnom poistení zodpovednosti za škodu spôsobenú prevádzkou motorového vozidla. Pri určení výšky poistného je poisťovňa povinná zohľadniť predchádzajúci bezškodový priebeh pri určovaní poistného a to zľavou. Ak poistený spôsobil škodu je poisťovňa povinná zohľadniť to prirážkou k poistnému.

Na Slovensku sa bonus malus systém začal uplatňovať v plnom rozsahu až 1. apríla 2015. Slovensko sa ako posledné zaradilo k európskym krajinám, ktoré bonus malus využíva.

Slovenský trh v oblasti povinného zmluvného poistenia v súčasnosti zaznamenáva silnú konkurenciu. Zákazníci si môžu vybrať poisťovňu, ktorá im ponúkne najlepšie podmienky. Poisťovne sa snažia vytvoriť čo najhomogénnejší poistný kmeň a stanoviť poistné, ktoré bude adekvátne zodpovedať skutočnému rizikovému profilu každého poistenca. To znamená jeho aktuálnu škodovú frekvenciu. Bonus malus systém slúži aj k motivácii vodičov k bezškodovosti.

Bonus malus sa skladá z určitých tried (stupňov). Tieto triedy hovoria o relatívnom poistnom v danej triede. Poistený sa každý rok zaradí do jednej triedy podľa toho aký bol jeho škodový priebeh počas predchádzajúceho roka. Prechody (prestupy) do jednotlivých tried sa riadia pravidlami, ktoré si stanovuje každá poisťovňa sama. Poistenec tak platí poistné podľa toho v akej triede sa nachádza.

Pri bonus malus systéme je potrebné definovať niekoľko pravidiel. Hlavným zdrojom tejto časti bolo (Fecenko 2012). Dôležité pri realizácii sú triedy bonus malus systému podľa ktorých sa určuje zrážka/prirážka k poistnému a pravidlá prechodu medzi jednotlivými triedami.

Každý bonus malus má nasledujúce vlastnosti:

- Každý poistenec sa na začiatku obdobia zaradí do jednej z bonus malusovej triedy, podľa predchádzajúceho škodového priebehu.
- $H, H = 1, 2, 3 \dots$ je počet bonus malusových tried, ktoré má daná poisťovňa.
- Poistné P závisí od toho v ktorej bonus malusovej triede sa poistenec nachádza: $P = P(j), j = 1, 2, \dots, H$, pričom $P(j) \leq P(j + 1)$,

$$j = 1, 2, \dots, H - 1$$
- Priradenie poistenca do danej bonus malusovej triedy závisí od pravidiel poisťovne, ktoré sú presne stanovené v matici **R**

Každý poistenec sa na začiatku poistného obdobia zaradí do bonus malusovej triedy, ktorá sa odvíja od škodového priebehu v predchádzajúcom roku a od triedy v ktorej sa nachádzal.

Pravidlá prestupu poistenca medzi jednotlivými triedami i (čo je východisková trieda) a j (čo je koncová trieda) sú stanovené v matici prestupu **$R = (N_{ij})$** typu $H \times H$. Prvky N_{ij} majú takéto vlastnosti:

- $N_{ij} \subset \{0, 1, 2, 3, \dots\}, i, j = 1, 2, 3, \dots, H$,
- $\bigcup_{j=1}^H N_{ij} = \{0, 1, 2, 3, \dots\}, i, j = 1, 2, 3, \dots, H$,
- $N_{ij} \cap N_{ik} = \emptyset$, a platí pre každé $j \neq k, i, j, k = 1, 2, 3, \dots, H$,

kde N_{ij} je množina, ktorej prvky určujú počet poistných udalostí v predchádzajúcom období. Poistenec, ktorý je v triede i prejde do triedy j , ak sa v predchádzajúcom období počet poistných udalostí rovná niektorému prvku z N_{ij} . Ak $N_{ij} = \emptyset$, tak sa prechod nedá uskutočniť. V matici prestupu takúto možnosť označíme bodkou.

Pre i -tu triedu, kde $i = 1, 2, 3, \dots, H$, výška poistného v danej triede $P(i)$ sa určuje v percentách zo základného poistného a platí že $P(i) \leq P(j)$ pre $i < j$. Z toho vieme usúdiť, že trieda i je výhodnejšia ako trieda j . Pravidlá prestupu môžu byť vyjadrené rôzne. Napríklad, ak poistenec nemal poistnú udalosť v predchádzajúcom období, postúpi o jednu triedu vyššie, čo znamená že dostane zľavu na poistnom. Ak sa poistenec nachádza

v najvyššej bonus malusovej triede, tak v nej zotrúva. Poistenec za každú poistnú udalosť klesne o dve triedy nižšie.

Každá poisťovňa má všetky informácie o bonus malus systéme zakotvené vo svojich všeobecných zmluvných podmienkach.

Prvky matice prechodu sú definované ako pravdepodobnosť

$$p_{ij}^{\lambda} = P^{\lambda}(S_{t+1} = j | S_t = i),$$

kde λ je škodová frekvencia nehodovosti. S_{t+1} a S_t sú náhodné premenné, ktoré vyjadrujú príslušnú bonus malusovú triedu v čase t a $t+1$.

Časový interval je konštantný a jeho dĺžka je 1 rok. Predpokladajme, že pravdepodobnosť $p_{ij}^{\lambda}(t)$ nezávisí od času t . Potom platí $p_{ij}^{\lambda}(t) = p_{ij}^{\lambda}$. Matica prechodu pre škodovú frekvenciu λ je $\mathbf{P}^{\lambda} = (p_{ij}^{\lambda})_H^H$. Pravdepodobnosť p_{ij}^{λ} vyjadruje, že poistenec ktorý sa nachádza v i -tej bonus malusovej triede sa bude o rok nachádzať v j -tej triede, so škodovou frekvenciou λ . Každý poistenec sa nachádza vždy iba v 1 triede. Z toho vyplýva, že súčet riadkov matice $\mathbf{P}^{\lambda}$ sa rovná 1.

Pre poistencov so škodovou frekvenciou v daných rokoch platí pre relatívnu početnosť ich očakávaného rozloženia do jednotlivých tried rekurentný vzťah

$$X^{\lambda}(k+1) = X^{\lambda}(k)\mathbf{P}^{\lambda}, k = 0, 1, 2, \dots,$$

kde $X^{\lambda}(0) = (0, 0, \dots, 1, \dots, 0)$ a je to jednotkový vektor. Jednotka je práve na l -tom mieste, ktoré určuje v ktorej bonus malusovej triede sa poistený nachádza. Súčet všetkých zložiek vektora $X^{\lambda}(k+1)$ je 1.

Pre bonus malus systémy, ktoré môžeme opísať pomocou Markovových reťazcov existuje vektor stacionárnych pravdepodobností. Tento vektor určuje pravdepodobnosť s akou sa budú poistenci vyskytovať v daných triedach v dlhodobom horizonte. Bonus malus systém musí spĺňať tieto tri vlastnosti:

- Množina stavov Markovovho reťazca je konečná.
- Do každej bonus malusovej triedy sa dá dostať z inej ľubovoľnej triedy. Táto vlastnosť sa volá ireducibilita.

- V bonus malus systéme sa nachádza najvýhodnejšia trieda, ak sa v nej poistenec nachádza a ak v nasledujúcom období nespôsobí škodu, tak ostáva v tejto triede. To znamená že stavy Markovovho reťazca sú aperiodické.

Vektor stacionárnych pravdepodobností X je definovaný $X^\lambda = \lim_{n \rightarrow \infty} X^\lambda(k)$ alebo rovnicou $X^\lambda = X^\lambda P^\lambda$ za predpokladu, že platí podmienka $x_1^\lambda + x_2^\lambda + \dots + x_n^\lambda = 1$, kde platí $X^\lambda = (x_1^\lambda + x_2^\lambda + \dots + x_n^\lambda)$.

Základné poistné $P = P_l$, kde l je trieda ktorá je vstupnou do bonus malus systému. Poistné P_i je poistné, ktoré platí poistenec v i -tej bonus malusovej triede. Potom priemerné poistné, ktoré dostane poisťovňa od poistencov so škodovou frekvenciou λ sa dá vyjadriť

$$\bar{P} = \sum_{i=1}^H x_i^\lambda P_i = \sum_{i=1}^H x_i^\lambda \frac{P(1)}{100} P.$$

Výskyt poistencov v jednotlivých bonus malusových triedach sa riadi Poissonovým rozdelením. Pomocou pravdepodobnostnej funkcie Poissonovho rozdelenia vytvoríme maticu prechodu

$$P(N = k) = e^{-\lambda} \frac{\lambda^k}{k!}, k = 0, 1, 2, \dots; \lambda > 0 \quad (4.1)$$

1.5. Zdravotné a nemocenské poistenie

Zdravotné a nemocenské poistenie má poistenca odľahčiť od nákladov pri chorobe alebo nákladov na liečbu a rozložiť ich na celý poistný kmeň. Zákonné zdravotné poistenie vo svete kryje služby na rozdiel od nemocenského poistenia, ktoré kryje liečebné náklady.

Na Slovensku máme zdravotné poistenie a nemocenské poistenie. Zdravotné poistenie hradí náklady na liečbu a prevenciu pred ochorením. Nemocenské poistenie na rozdiel od zdravotného poistenia hradí stratu príjmu pri ochorení. Vo veľa krajinách je súčasťou zdravotného poistenia aj nemocenské poistenie. Používa sa iba jeden termín v prípade, ak sa tieto dva druhy poistenia nerozlišujú – zdravotné poistenie.

V krajinách, ktoré majú vyspelý poistný trh je viac rozvinuté životné poistenie. Nemocenské poistenie je zvyčajne súčasťou poistenia na úmrtie alebo poistenia na dožitie. Toto poistenie poskytuje komplexnejšiu ochranu poistenca. Poistenie sa vypláca buď žijúcemu poistencovi alebo pozostalým po poistencovi.

V zdravotnom a nemocenskom poistení bolo zostrojených veľa výpočtových modelov. Hlavným cieľom bolo nájsť spoľahlivé metódy na výpočet hodnoverných výsledkov. Najdôležitejšie je pre poisťovňu poznať výšku poistného, ktoré má platiť poistenec.

Každá poisťovňa má vlastné podmienky pri ktorých vypláca poistencom poistnú sumu. Tieto podmienky sú zakomponované do modelu už pri jeho vytváraní. Ochrana pred podvodmi je veľmi dôležitá pre poisťovne. Jeden zo spôsobov ochrany je tzv. čakacia doba. Používa sa hlavne pri nových poistencoch. Je to zvyčajne 1 rok.

2. Ciel práce

Hlavným cieľom práce je aplikácia Markovových reťazcov v oblasti aktuárstva. Ukážka v oblasti životného aj neživotného poistenia. Ďalším cieľom je definovať Markovové reťazce. Hlavným nástrojom na výpočet bude software R. V tomto software budeme demonštrovať balíčky, ktoré boli vytvorené pre Markovové reťazce. Práca obsahuje aj čiastkové ciele. V práci by mali byť opísané postupy pre tvorbu matíc pravdepodobnosti prechodu pomocou frekvencie škody a nagenetrovaných dát. To sa bude demonštrovať na dátach, ktoré sú z Talianska. Ďalším cieľom práce je ukážka výpočtu poistného pomocou simulácií Markovovho reťazca.

3. Metodika práce a metódy skúmania

V tejto časti sa oboznámime s riešením problematiky Markovových reťazcov pomocou software R.

Markovové reťazce predstavujú triedu stochastických procesov so širokým spektrom praktickej aplikácie. Konkrétne, diskretný čas Markovovho reťazca nám dovoľuje modelovať pravdepodobnosti prechodu medzi jednotlivými stavmi pomocou matice. Mnohé R balíčky sa zaoberajú modelmi, ktoré sú založené na Markovových reťazcoch :

- `msm` (Jackson 2011) spracováva viacstavové modely pre dlhodobé dáta
- `mcmcR` (Geyer a Johnson 2013) implementuje prístup Monte Carlo na Markovove reťazce
- `hmm` (Himmelman a www.linhi.com 2010) sa používa na skryté Markovové reťazce
- `mstate` (de Wreede, Fiocco a Putter 2011) sa používa pri analýze prežitia založenej na Markovovom reťazci (Spedicato 2018)

Prostrediu R chýbal jednoduchý balík, ktorý by zahŕňal široké množstvo aplikácií pre Markovové reťazce. Preto bol vytvorený balík `markovchain`. Tento balík si kladie za cieľ ponúknuť väčšiu flexibilitu pri manipulácii s Markovovým reťazcom. Balík `markovchain` je založený na nasledujúcich balíkoch:

- `expm` (Goulet, Dutang, Maechler, Firth, Shapira, Stadelmann, and expm-developers@lists.R-forge.R-project.org 2013) vykonáva maticové operácie
- `igraph` (Csardi a Nepusz 2006) slúži na vykresľovanie
- `matlab` (Roebuck 2011) obsahuje maticové funkcie a výpočty, ktoré napodobňujú prostredie v prostredí MATLAB (Spedicato 2018)

V nasledujúcich častiach budeme ukazovať štruktúru balíka a postupnú aplikáciu na trojstavový model. Budú to zjednodušené ukážky. Reálna aplikácia bude v časti výsledkov práce.

3.1. Vytváranie objektov typu `markovchain`

Ako prvý krok je treba načítať z knižnice príkaz:

```
library("markovchain")
```

Každý prvok triedy `markovchain` sa skladá z nasledujúcich vstupov:

1. `states`: znakový vektor, ktorý uvádza stavy, pre ktoré sú pravdepodobnosti prechodu definované
2. `byrow`: logický prvok označujúci, či sú pravdepodobnosti zobrazené v riadkoch alebo v stĺpcoch
3. `transitionMatrix`: pravdepodobnosti v prechodovej matici
4. `name`: voliteľný znakový parameter prvkov na pomenovanie (Spedicato 2018)

Objekt `markovchain` môžeme vytvoriť viacerými spôsobmi. Prvý spôsob je zdĺhavejší. Spočíva v tom, že jednotlivé vstupy sa samostatne zadefinujú a potom sa zložia do Markovovho reťazca.

```
stav<-c("1. stav", "2. stav", "3. stav")
byRow <- TRUE
mcStav<-matrix(data = c(0.60, 0.3, 0.1,
                        0.3, 0.4, 0.3,
                        0.2, 0.45, 0.35),
               byrow = byRow, nrow = 3,
               dimnames = list(stav,stav),)
mcStav
```

(6.1)

Druhá možnosť je kratšia. Do celého príkazu sa zadefinujú všetky vstupy.

```
mcStav <- new("markovchain", states = c("1. stav", "2. stav", "3. stav"),
             transitionMatrix = matrix(data = c(0.60, 0.3, 0.1,
                                                0.3, 0.4, 0.3,
                                                0.2, 0.45, 0.35),
                                       byrow = byRow, nrow = 3),
             name = " Trojstavový Markovov reťazec ")
```

Výstup oboch možností je rovnaký objekt typu `markovchain`.

```
Trojstavový Markovov reťazec
A 3 - dimensional discrete Markov Chain defined by the following states:
1. stav, 2. stav, 3. stav
The transition matrix (by rows) is defined as follows:
  1. stav 2. stav 3. stav
1. stav   0.6   0.30  0.10
2. stav   0.3   0.40  0.30
3. stav   0.2   0.45  0.35
```

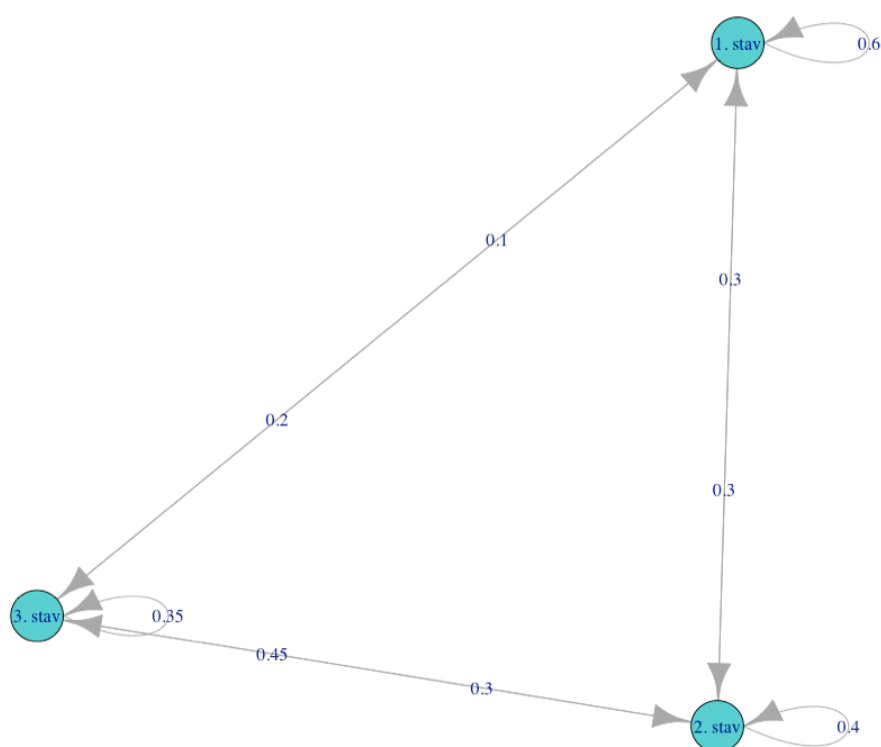
Vo výstupe môžeme vidieť všetky informácie o Markovovom reťazci. Je v ňom zobrazená matica pravdepodobnosti prechodu Markovovho reťazca, stavy, dimenzia a názov reťazca (Spedicato 2017).

3.2. Vykresľovanie objektov typu `markovchain`

Prechodovú maticu objektu `markovchain` môžeme zobrazit' pomocou dvoch metód. Prvá metóda zobrazenia je základný diagram pravdepodobností prechodu.

```
plot(mcStavy)
```

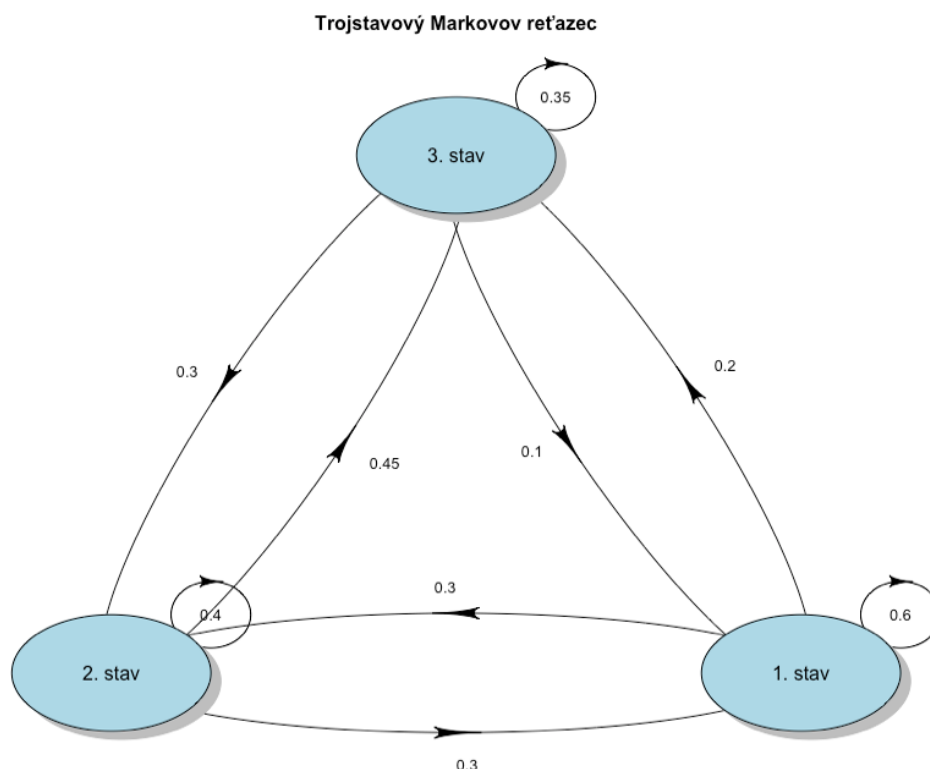
Výstupom tohoto príkazu je Obr. č. 1. V kruhoch sú zapísané názvy stavov. Číslo pri kruhoch je pravdepodobnosť zotrvania v danom stave. Číslo bližšie k stavu vyjadruje pravdepodobnosť prechodu do daného stavu. V tomto prípade si nemôžeme vybrať farbu ani veľkosť. Farby grafu sú náhodné. Je veľmi neprehľadný, preto sa na vykreslenie používa druhá možnosť (Spedicato 2017).



Obr. č. 1: Jednoduchý graf Markovovho reťazca

Druhá metóda ma rozšírený zápis. Môžeme si vybrať farbu, veľkosť písma, hrúbku písma a aj hrúbku čiary. Táto možnosť je prehľadnejšia ako môžeme vidieť na Obr. č. 2. V tomto prípade je potrebné, aby matica prechodu bola samostatne definovaná a nie ako súčasť Markovovho reťazca.

```
matica <- as(mcStavy,"matrix")
plotmat(matica ,lwd = 1, box.lwd = 1,
cex.txt = 0.8,box.size = 0.1,box.type = "circle",
box.prop = 0.7,box.col = "light blue",arr.length=.4,
arr.width=.1,self.cex = .4,main = "Trojstavový Markovov reťazec")
```

(6.2)


Obr. č. 2: Prehľadnejší graf Markovovho reťazca

Ako pri predchádzajúcom vykreslení v elipse sa nachádza názov stavu. Číslo pri elipse je pravdepodobnosť zotrvania v danom stave. Šípka ukazuje k stavu z ktorého sa vychádza. Číslo pri tejto šípke je pravdepodobnosť prechodu do stavu, ktorý je spojený čiarou (Spedicato 2017).

3.3. Pravdepodobnostné operácie s objektami typu `markovchain`

Balík `markovchain` obsahuje funkcie na analýzu Markovových reťazcov s pravdepodobnostnou perspektívou. Balík napríklad poskytuje metódy na nájdenie vektora absolútnych pravdepodobností a identifikáciu absorpčných a prechodových stavov. Mnohé z týchto metód pochádzajú z MATLAB, ktoré boli prenesené do R. Tab. č. 1 ukazuje niektoré metódy, ktoré môžeme použiť na pravdepodobnostné analýzy Markovových reťazcov (Spedicato 2017).

Tab. č. 1: Pravdepodobnostné operácie

metóda	výstup
<code>absorbingStates</code>	absorpčné stavy prechodovej matice, ak existujú
<code>steadyStates</code>	vektor (y) ustáleného stavu (ov) vo forme matice. (6.3)
<code>recurrentClasses</code>	zoznam opakujúcich sa tried
<code>summary</code>	Súhrn Markovovho reťazca
<code>transientStates</code>	prechodné stavy prechodovej matice, ak existujú

Výstupy z daných príkazov vyzerajú nasledovne.

```
> absorbingStates(mcStavy)
character(0)
> steadyStates(mcStavy)
      1. stav  2. stav  3. stav
[1,] 0.3953488 0.372093 0.2325581
> recurrentClasses(mcStavy)
[[1]]
[1] "1. stav" "2. stav" "3. stav"
> summary(mcStavy)
Trojstavový Markovov reťazec   Markov chain that is composed by:
Closed classes:
1. stav 2. stav 3. stav
Recurrent classes:
{1. stav, 2. stav, 3. stav}
Transient classes:
NONE
The Markov chain is irreducible
The absorbing states are: NONE
```

3.4. Štatistická analýza a simulácie

V nasledujúcej časti sa budeme venovať štatistickým analýzám a simuláciám, ktoré sú implementované v rámci balíka `markovchain`, a ktoré pomáhajú prispôbiť, simulovať alebo predpovedať Markovov reťazec. Príkazy sú vedené v Tab. č. 2 (Spedicato 2017).

Tab. č. 2: Štatistická analýza

metóda	výstup
<code>markovchainFit</code>	Funkcia vráti Markovov reťazca pre danú sekvenciu
<code>predict</code>	Metóda výpočtu predpovedí z objektu typu <code>markovchain</code>
<code>rmarkovchain</code>	Funkcia na vzorkovanie objektov typu <code>markovchain</code>

Simulácie náhodnej sekvencie z Markovovho reťazca sú vďaka tejto funkcii celkom jednoduché. Nasledujúci kód vygeneruje 100 možných stavov podľa nášho trojstavového modelu. V tejto simulácii si môžeme vybrať aj stav, ktorý je ako počiatočný. My sme si za počiatočný stav zvolili 1. stav. Je to označené v premennej s názvom `t0`.

```
vigerenovaneStavy <- rmarkovchain(n = 100, object = mcStavy, t0 = "1. stav")  
vigerenovaneStavy
```

(6.4)

Výstupom príkazov je 100 možných prechodov zo stavov 1. stav, 2. stav a 3. stav. Počiatočná podmienka bola zvolená tak, že východiskový stav je 1. stav. Je to rad po sebe idúcich stavov, ktoré boli vygenerované na základe zadanej matice prechodu.

```
[1] "3. stav" "3. stav" "2. stav" "1. stav" "2. stav" "2. stav" "2. stav" "2. stav" "2. stav"  
[9] "1. stav" "1. stav" "2. stav" "3. stav" "2. stav" "1. stav" "1. stav" "1. stav" "1. stav"  
[17] "1. stav" "1. stav" "1. stav" "2. stav" "1. stav" "1. stav" "1. stav" "2. stav" "2. stav"  
[25] "2. stav" "3. stav" "2. stav" "2. stav" "3. stav" "2. stav" "2. stav" "2. stav" "2. stav"  
[33] "2. stav" "2. stav" "2. stav" "2. stav" "2. stav" "3. stav" "1. stav" "2. stav" "2. stav"  
[41] "2. stav" "1. stav" "1. stav" "1. stav" "1. stav" "1. stav" "1. stav" "2. stav" "2. stav"  
[49] "3. stav" "2. stav" "2. stav" "2. stav" "1. stav" "1. stav" "1. stav" "2. stav" "2. stav"  
[57] "3. stav" "2. stav" "1. stav" "1. stav" "1. stav" "2. stav" "3. stav" "3. stav" "3. stav"  
[65] "2. stav" "3. stav" "2. stav" "2. stav" "1. stav" "1. stav" "1. stav" "2. stav" "2. stav"  
[73] "3. stav" "3. stav" "1. stav" "1. stav" "1. stav" "1. stav" "3. stav" "3. stav" "3. stav"  
[81] "2. stav" "2. stav" "1. stav" "1. stav" "2. stav" "1. stav" "1. stav" "3. stav" "3. stav"  
[89] "3. stav" "3. stav" "1. stav" "3. stav" "2. stav" "2. stav" "1. stav" "1. stav" "1. stav"  
[97] "1. stav" "1. stav" "1. stav" "1. stav"
```

Časovo homogénny Markovov reťazec môže byť vytvorený aj z údajov. Boli použité metódy z kapitoly 1.3. Modelovanie matice pravdepodobnosti prechodu v Markovových reťazcoch. Rovnica (3.4) ukazuje odhad maximálnej pravdepodobnosti. Pomocou softvéru R vieme vytvoriť maticu pravdepodobností prechodu aj z dát ktoré poznáme. Pre túto ukážku použijeme dáta, ktoré sme si vygenerovali v predchádzajúcom kroku. Nasledujúcim kódom vytvoríme Markovov reťazec, v ktorom bude zahrnutá aj matica pravdepodobností prechodu (Spedicato 2017).

```
odhadStavy <- markovchainFit(data = vigerenovaneStavy, method = "mle",
                             name = "Trojstavový Markovov reťazec")

odhadStavy$estimate
```

Výstupom je Markovov reťazec, v ktorom je zakomponovaná aj matica pravdepodobností prechodu

```
Trojstavový Markovov reťazec
A 3 - dimensional discrete Markov Chain defined by the following states:
1. stav, 2. stav, 3. stav
The transition matrix (by rows) is defined as follows:
      1. stav  2. stav  3. stav
1. stav 0.5813953 0.3720930 0.04651163
2. stav 0.4324324 0.2972973 0.27027027
3. stav 0.1052632 0.5263158 0.36842105
```

Ako môžeme vidieť na matici, ktorú sme zadali a na matici z vygenerovaných dát, matice prechodu sa líšia. Ale ak by sa uskutočnil väčší počet simulácií, matice by boli takmer totožné.

Ak by sme chceli predpovedať stavy niekoľko krokov dopredu, vieme to odhadnúť pomocou matice pravdepodobností prechodu. Táto metóda sa dá použiť pre homogénny aj nehomogénny Markovov reťazec.

Na predpovedanie nasledujúcich štyroch stavov použijeme maticu pravdepodobnosti prechodu, ktorú sme odhadli pomocou vygenerovaných dát. V tomto prípade môžeme zadať do príkazu predchádzajúce stavy (Spedicato 2017).

```
predict(object = odhadStavy$estimate, newdata = c("2. stav", "3. stav"), n.ahead = 4)
```

Výstupom sú štyri stavy, ktoré budú nasledovať.

```
[1] "2. stav" "1. stav" "1. stav" "1. stav"
```

4. Výsledky práce a diskusia

V tejto časti budeme aplikovať poznatky z predchádzajúcich kapitol na konkrétne prípady. V oblasti neživotného poistenia ich budeme aplikovať na bonus malus systém. Ukážeme vytáranie matice pravdepodobností prechodu a výpočet stacionárneho vektora. V životnom a neživotnom poistení bude naša aplikácia na výpočet poistného. V diskusií následne zhrnieme naše výsledky.

4.1. Aplikácia poznatkov na bonus malus systém

Teraz aplikujeme poznatky na konkrétny bonus malus systém. Pri výpočtoch budeme používať software R. Bonus malus systém, ktorý budeme analyzovať je iba príklad no dá sa aplikovať na akúkoľvek poisťovňu.

Rozhodná doba je vo všeobecnosti neprerušená doba trvania poistenia, ktorá sa počíta v ukončených celých mesiacoch, pričom je znížená ak nastane poistná udalosť. V našom prípade je vstupná trieda 8 (B0). Vo vstupnej triede sa platí poistné vo výške 100 %. Systém má 7 bonusových tried a 4 malusové triedy. Ak poistenec nemal v predchádzajúcom roku žiadnu poistnú udalosť postúpi o 1 triedu nahor, maximálne však do prvej bonusovej triedy. Ak sa nachádza v najvyššej triede, zotrvá v nej. Ak poistenec nahlási škodu, posunie sa o 2 triedy nadol. V takomto prípade sa rozhodná doba zníži o 24 mesiacov. Najnižšia je 12. trieda (M4), v ktorej sa platí poistné vo výške 180 %. Výška poistného v jednotlivých bonus malusových triedach je zobrazená v Tab. č. 3 (Fecenk 2012).

Tab. č. 3: Výška poistného v jednotlivých bonus malus triedach

bonus malus trieda	Stupeň bonusu a malusu	Výška poistného
1	B7	50%
2	B6	60%
3	B5	70%
4	B4	80%
5	B3	85%
6	B2	90%
7	B1	95%
8	B0	100%

9	M1	110%
10	M2	130%
11	M3	150%
12	M4	180%

Matica pravidiel prestupu je v našom prípade zobrazená na Obr. č. 3. V tejto matici je vidieť, ktoré stavy sú dosiahnuteľné a ktoré nie. Dosiahnuteľný stav je taký stav do ktorého môže poistenec prejsť. Čísla vyjadrujú počet poistných udalostí.

	1	2	3	4	5	6	7	8	9	10	11	12
1	{0}	.	{1}	.	{2}	.	{3}	.	{4}	.	{5}	{6, 7, ...}
2	{0}	.	.	{1}	.	{2}	.	{3}	.	{4}	.	{5, 6, ...}
3	.	{0}	.	.	{1}	.	{2}	.	{3}	.	{4}	{5, 6, ...}
4	.	.	{0}	.	.	{1}	.	{2}	.	{3}	.	{4, 5, ...}
5	.	.	.	{0}	.	.	{1}	.	{2}	.	{3}	{4, 5, ...}
6	.	.	.	.	{0}	.	.	{1}	.	{2}	.	{3, 4, ...}
7	.	.	.	.	.	{0}	.	.	{1}	.	{2}	{3, 4, ...}
8	.	.	.	.	.	.	{0}	.	.	{1}	.	{2, 3, ...}
9	.	.	.	.	.	.	.	{0}	.	.	{1}	{2, 3, ...}
10	.	.	.	.	.	.	.	.	{0}	.	.	{1, 2, ...}
11	.	.	.	.	.	.	.	.	.	{0}	.	{1, 2, ...}
12	.	.	.	.	.	.	.	.	.	.	{0}	{1, 2, ...}

Obr. č. 3: Matica pravidiel prestupu

Napríklad, ak by sa poistenec nachádzal v 5. bonus malusovej triede a za predchádzajúci rok by nenahlásil žiadnu poistnú udalosť, posunul by sa o jednu triedu vyššie a teda do 4. triedy. Naopak, ak by nahlásil jednu poistnú udalosť, posunul by sa o dve triedy nižšie, teda do 7 triedy. Ak by nahlásil dve poistné udalosti, posunul by sa o štyri triedy nižšie, teda do 9. triedy. Ak by poistenec spôsobil viac ako štyri poistné udalosti, dostal by sa do poslednej 12. triedy (Fecenko 2012).

Predpokladáme, že škodová frekvencia λ je 0,1. To znamená že priemerný Slovák, ktorý má uzatvorené PZP nahlási škodu raz za desať rokov. Výskyt poistencov v jednotlivých bonus malusových triedach sa riadi Poissonovým rozdelením. Parametrom rozdelenia je naša škodová frekvencia $\lambda = 0,1$ (Fecenko 2012).

Pomocou pravdepodobnostnej funkcie Poissonovho rozdelenia, ktorá je uvedená vo vzťahu (4.1) vytvoríme Markovov reťazec. Reťazec je vytvorený analogicky podľa (6.1).

Celý kód Markovoho reťazca je zadáný vo funkcii, v ktorej je parameter škodová frekvencia (Spedicato 2018).

```
BMretazec12 <- function(lambda)
{
  bmM12 <- matrix(0, 12, 12)

  bmM12[1, 1] <- dpois(x = 0, lambda)
  bmM12[1, 3] <- dpois(x = 1, lambda)
  ...
  bmM12[12, 11] <- dpois(x = 0, lambda)
  bmM12[12, 12] <- 1 - ppois(q = 0, lambda)

  stavyBM12 <- as.character(1:12)
  BM12 <- new("markovchain", transitionMatrix = bmM12,
             states = stavyBM12, name = "BM Matica") ##vytvorenie markovovho retazca matice
  pravdepodobosti
  return(BM12) ##funkcia vrati maticu
}

BMM12 <- BMretazec12(0.1) ###vytvorenie matice s daným parametom
BMM12
```

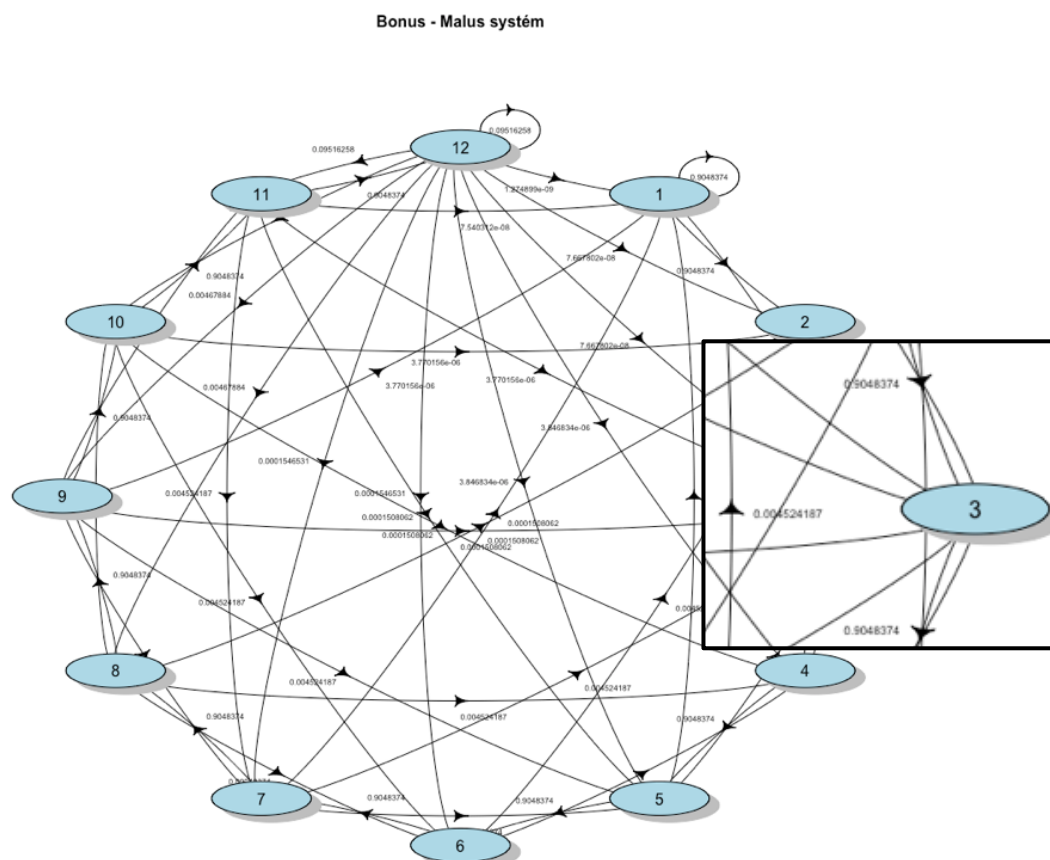
Logika funkcie spočíva v tom, že vytvoríme nulovú maticu s rozmermi 12 x12. Parametre matice závisia od počtu bonus malusových tried. Následne sa podľa súradníc prvku priradí do nulovej matice pravdepodobnosť podľa toho, koľko bolo poistných udalostí. Takto sa vyplnia všetky prvky matice. V prípade, že tento stav nie je dosiahnuteľný, zostáva v matici 0. Následne sa vytvoria názvy stavov. V našom prípade sú stavy označené číslami 1 až 12. Posledný krok funkcie je vytvorenie Markovovho reťazca so všetkými vstupmi, ktoré má mať. Následne vyvoláme funkciu, v ktorej určíme škodovú frekvenciu. V našom prípade je to 0,01.

Výstupom kódu je Markovov reťazec s maticou pravdepodobností prechodu pre náš bonus malus systém.

```
BM Matica
A 12 - dimensional discrete Markov Chain defined by the following states:
1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12
The transition matrix (by rows) is defined as follows:
      1      2      3      4      5      6      7      8      9      10     11     12
1  0.9048374 0.0000000 0.09048374 0.00000000 0.004524187 0.000000000 0.0001508062 0.000000000 3.770156e-06 0.000000e+00 7.540312e-08 1.274899e-09
2  0.9048374 0.0000000 0.00000000 0.09048374 0.000000000 0.004524187 0.000000000 0.0001508062 0.000000e+00 3.770156e-06 0.000000e+00 7.667802e-08
3  0.0000000 0.9048374 0.00000000 0.00000000 0.090483742 0.000000000 0.0045241871 0.000000000 1.508062e-04 0.000000e+00 3.770156e-06 7.667802e-08
4  0.0000000 0.0000000 0.90483742 0.00000000 0.000000000 0.090483742 0.000000000 0.0045241871 0.000000e+00 1.508062e-04 0.000000e+00 3.846834e-06
5  0.0000000 0.0000000 0.00000000 0.90483742 0.000000000 0.000000000 0.0904837418 0.000000000 4.524187e-03 0.000000e+00 1.508062e-04 3.846834e-06
6  0.0000000 0.0000000 0.00000000 0.00000000 0.904837418 0.000000000 0.000000000 0.0904837418 0.000000e+00 4.524187e-03 0.000000e+00 1.546531e-04
7  0.0000000 0.0000000 0.00000000 0.00000000 0.000000000 0.904837418 0.000000000 0.000000000 9.048374e-02 0.000000e+00 4.524187e-03 1.546531e-04
8  0.0000000 0.0000000 0.00000000 0.00000000 0.000000000 0.000000000 0.9048374180 0.000000000 0.000000e+00 9.048374e-02 0.000000e+00 4.678840e-03
9  0.0000000 0.0000000 0.00000000 0.00000000 0.000000000 0.000000000 0.000000000 0.9048374180 0.000000e+00 0.000000e+00 9.048374e-02 4.678840e-03
10 0.0000000 0.0000000 0.00000000 0.00000000 0.000000000 0.000000000 0.000000000 0.000000000 9.048374e-01 0.000000e+00 0.000000e+00 9.516258e-02
11 0.0000000 0.0000000 0.00000000 0.00000000 0.000000000 0.000000000 0.000000000 0.000000000 0.000000e+00 9.048374e-01 0.000000e+00 9.516258e-02
12 0.0000000 0.0000000 0.00000000 0.00000000 0.000000000 0.000000000 0.000000000 0.000000000 0.000000e+00 0.000000e+00 9.048374e-01 9.516258e-02
```

Celý kód na vytvorenie Markovovho reťazca nájdeme v prílohe č.1.

Vykreslenie bonus malus systému je analogické podľa (6.2). Celý kód nájdeme v prílohe č. 2.



Obr. č. 4: Vizualizácia bonus malus systému

Číslo v elipse je názov bonus malusovej triedy. Čiary spájajú stavy medzi ktorými je možný prechod. Šípka ukazuje na stav z ktorého sa vychádza.

Vektor počiatkových pravdepodobností $X^\lambda(0)$ v našom prípade vyzerá

$$X^\lambda(0) = (0,0,0,0,0,0,0,1,0,0,0,0).$$

Ako môžeme vidieť náš vstupný stav je 8. bonus malusová trieda. Do tejto triedy je zaradený každý nový poistenec. Vektor absolútnych pravdepodobností vypočítame pomocou (6.3). Kód na výpočet vyzerá nasledovne.

```
as.numeric(steadyStates(BMM12))
```

Výstupom je vektor absolútnych pravdepodobností. V našom prípade vektor vyzerá nasledovne.

```
[1] 0.7789946 0.08192758 0.09054398 0.02216711 0.01630569 0.0050712 0.00297819  
0.001078292 0.0005600 0.00022130 0.000107 0.00004459
```

Tento vektor nám ukazuje v akom pomere sa poistenci nachádzajú v jednotlivých triedach bonus malus systému. Toto platí, ak predpokladáme, že systém je ustálený. Systém je ustálený, ak sa nemení v závislosti od času. V našom prípade systém nie je stacionárny ani po 20 rokoch. Najviac poistencov sa nachádza v prvej bonus malusovej triede. V tejto triede platí poistenec 50 % zo základného poistného.

Pre poisťovňu je dôležité vedieť priemerné poistné, ktoré budú poistenci platiť. Toto poistné si môžeme vypočítať. Vytvoríme si vektor v ktorom budú percentá poistného v daných bonus malusových triedach, ktoré sú zaznamenané v tab. č. 3. Následne vynásobíme s vektorom absolútnych pravdepodobností (Spedicato 2018).

```
P<-c(0.5, 0.6, 0.7, 0.8, 0.85, 0.9, 0.95, 1, 1.1, 1.3, 1.5, 1.8)
sum(as.numeric(steadyStates(BMM12)) * P )
```

Výstupom je priemerné poistné ktoré platí každý poistenec za predpokladu ustáleného stavu.

```
[1] 0.5432449
```

Priemerné poistné je 54,32 % z poistného, ktoré bolo stanovené pri vstupe do bonus malusového systému. Poisťovňa môže tento poznatok použiť na stanovenie výšky poistného tak, aby vytvárala v nasledujúcom období zisk.

4.2. Aplikácia v oblasti životného a nemocenského poistenia

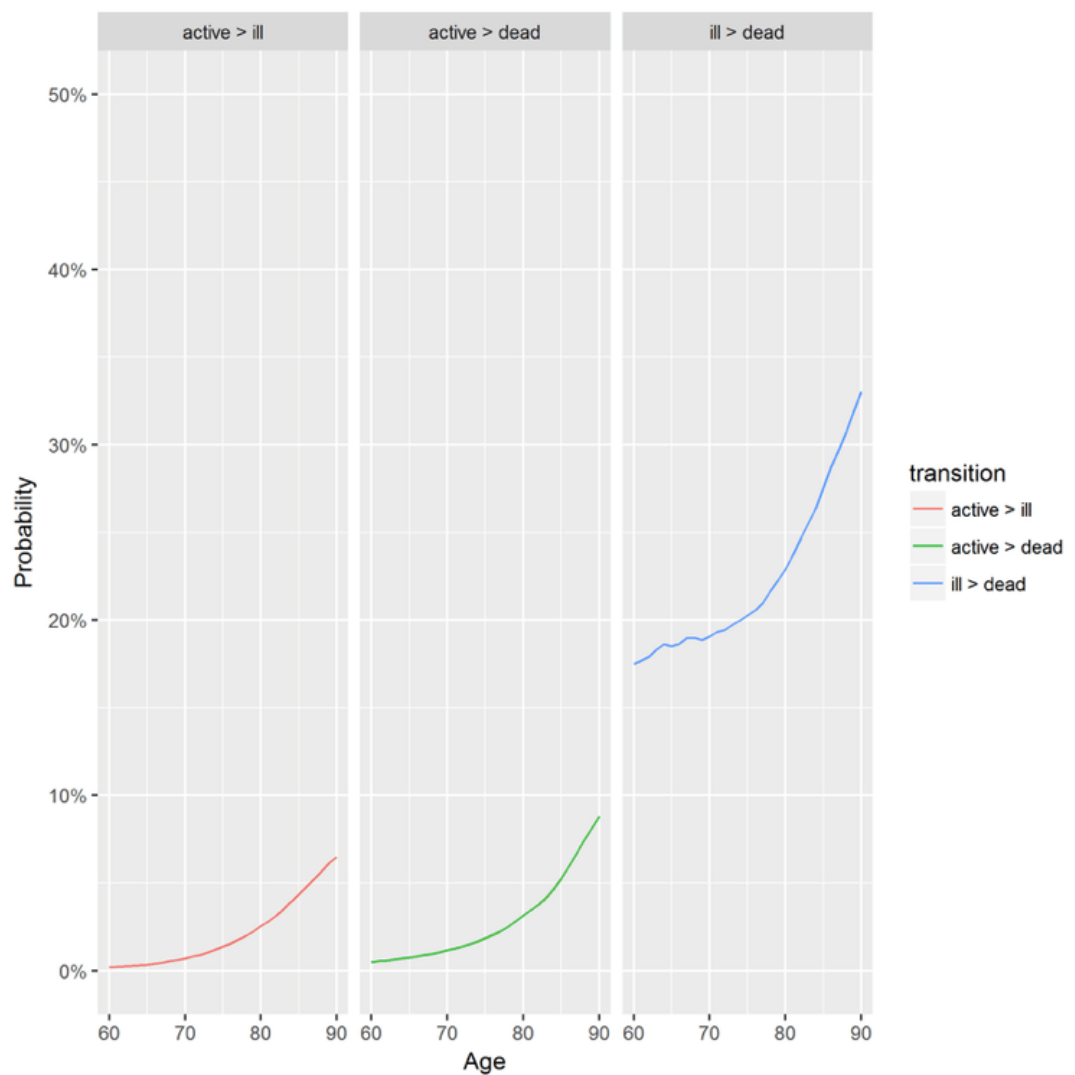
V tejto časti sa budeme zaoberať výpočtom poistného. Naše dáta pochádzajú z Talianska. Dáta sú zaznamenané za rok 2016. Je v nich zahrnutá iba mužská populácia. Sú v nich zaznamenané pravdepodobnosti prechodu medzi stavmi pre populáciu vo veku od 20 do 120 rokov. Dáta sa zaoberajú nesebestačnosťou. Čo je stav spôsobený chorobou alebo úrazom. Človek je nesebestačný, ak nedokáže vykonávať niektorú z činností ako je neschopnosť prijímania stravy, umývania sa, výkonu fyziologickej potreby, obliekania sa, strata mobility a schopnosti pohybu. Naše dáta predpokladajú, že osoba ktorá je nesebestačná sa už do stavu zdravý nevráti. Pre jednoduchosť o nesebestačnosti budeme hovoriť ako o chorobe. Ako prvý krok načítame dáta do R (Spedicato, 2018).

```
haltcDemoPath<-system.file("extdata", "ltdItaData.txt", package = "markovchain")
ltdcDemo<-read.table(file = ltdcDemoPath, header=TRUE, sep=";", dec = ".")
head(ltdcDemo)
```

Výstupom je tabuľka, v ktorej sú zaznamenané dáta o populácii.

	age	pAD	pID	pAI	pAA
1	20	0.0004616002	0.01083364	0.0001762467	0.9993622
2	21	0.0004824888	0.01079719	0.0001710577	0.9993465
3	22	0.0004949938	0.01177076	0.0001592333	0.9993458
4	23	0.0005042935	0.01159394	0.0001605731	0.9993351
5	24	0.0005074193	0.01260574	0.0001606504	0.9993319
6	25	0.0005154267	0.01526364	0.0001643603	0.9993202

V tabuľke vidíme vek osoby (age), pravdepodobnosť, že osoba zo stavu zdravý prejde do stavu mŕtvy (pAD), pravdepodobnosť, že osoba zo stavu chorý prejde do stavu mŕtvy (pID), pravdepodobnosť, že osoba zo stavu zdravý prejde do stavu chorý (pAI) a pravdepodobnosť, že osoba zotrvá v stave zdravý (pAA). Na obr. č. 5 vidíme ako pravdepodobnosti prechodu rastú.



Obr. č. 5: Pravdepodobnosti prechodu zo stavov pre mužov z Talianska za rok 2016

Aby sme z dát mohli vytvoriť Markovové reťazce potrebujeme doplniť chýbajúce pravdepodobnosti prechodu. Tým že osoba už nemôže vyzdraviť, tak pravdepodobnosť prechodu zo stavu chorý do stavu zdravý (pIA) bude 0. Ak človek zomrie, tak z tohto stavu sa už nemôže nikam posunúť. To znamená, že pravdepodobnosť zotrvania v stave mŕtvy (pDD) je 1. Tento stav nazývame absorpčný. Pravdepodobnosti prechodu zo stavu mŕtvy do stavu chorý (pDI) a do stavu zdravý (pDA) sú 0. Teraz doplníme dáta o tieto pravdepodobnosti.

```
ltcDemo<-transform(ltcDemo,
                  pIA=0,
                  pII=1-pID,
                  pDD=1,
                  pDA=0,
                  pDI=0)
head(ltcDemo)
```

Výstupom je tabuľka, ktorá je doplnená o pravdepodobnosti, ktoré sme zadefinovali.

	age	pAD	pID	pAI	pAA	pIA	pII	pDD	pDA	pDI
1	20	0.0004616002	0.01083364	0.0001762467	0.9993622	0	0.9891664	1	0	0
2	21	0.0004824888	0.01079719	0.0001710577	0.9993465	0	0.9892028	1	0	0
3	22	0.0004949938	0.01177076	0.0001592333	0.9993458	0	0.9882292	1	0	0
4	23	0.0005042935	0.01159394	0.0001605731	0.9993351	0	0.9884061	1	0	0
5	24	0.0005074193	0.01260574	0.0001606504	0.9993319	0	0.9873943	1	0	0
6	25	0.0005154267	0.01526364	0.0001643603	0.9993202	0	0.9847364	1	0	0

Teraz vytvoríme funkciu, ktorej premennou bude vek osoby. Táto funkcia nám vráti Markovov reťazec pre osobu vo veku, ktorý sme zadali. My budeme predpokladať, že táto osoba bude vo veku 40 rokov. Sú tam zaznamenané pravdepodobnosti prechodu osoby v zadanom veku. Funkcia bola vytvorená analogicky podľa (6.1).

```
stavyPo<-c("A","I","D") ##nazov stavou
MCvek<-function(age)
{
  prechodVek<-ltcDemo[ltcDemo$age==age,]
  prechodM<-matrix(0, nrow=3,ncol = 3,
                  dimnames = list(stavyPo,stavyPo))
  prechodM[1,1]<-prechodVek$pAA[]
  prechodM[1,2]<-prechodVek$pAI[]
  prechodM[1,3]<-prechodVek$pAD[]
  prechodM[2,2]<-prechodVek$pII[]
  prechodM[2,3]<-prechodVek$pID[]
  prechodM[3,3]<-1
  PoM<-new("markovchain",transitionMatrix=prechodM,
          states=stavyPo,
```

```

        name=paste("vek",age,"Prechodova matica"))
    return(PoM)
}
a<-40
MCvek(a)

```

Výstupom je Markovov reťazec pre osobu vo 40 rokoch.

```

vek 40 Prechodova matica
A 3 - dimensional discrete Markov Chain defined by the following states:
A, I, D
The transition matrix (by rows) is defined as follows:
      A      I      D
A 0.9988761 0.0002646749 0.0008592072
I 0.0000000 0.9414875933 0.0585124067
D 0.0000000 0.0000000000 1.0000000000

```

V ďalšom kroku vytvoríme funkciu, ktorá nám vytvorí všetky Markovové reťazce od zadaného veku po konečný vek, ktorý je v našom prípade 120 rokov. Tento objekt sa nazýva `list of Markov chain(s)`.

```

kompletnaPrechodM<-function(age)
{VekovaPostupnost<-seq(from=age, to=120)
  k=1
  zoznam=list()
  for ( j in VekovaPostupnost)
  {
    MCveki<-MCvek(age = j)
    zoznam[[k]]<-MCveki
    k=k+1
  }
  MCPo<-new("markovchainList", markovchains = zoznam,
            name = paste("PrechodOdVeku", age, sep = ""))

  return(MCPo)}
kompletnaPrechodM(a)

```

Výstupom je objekt, ktorý obsahuje všetky matice pravdepodobností prechodu od zadaného veku po koniec úmrtnostných tabuliek.

```

PrechodOdVeku40 list of Markov chain(s)
Markovchain 1
vek 40 Prechodova matica
A 3 - dimensional discrete Markov Chain defined by the following states:
A, I, D
The transition matrix (by rows) is defined as follows:
      A      I      D
A 0.9988761 0.0002646749 0.0008592072
I 0.0000000 0.9414875933 0.0585124067
D 0.0000000 0.0000000000 1.0000000000

```

```

Markovchain 2
vek 41 Prechodova matica
A 3 - dimensional discrete Markov Chain defined by the following states:
A, I, D
The transition matrix (by rows) is defined as follows:
      A      I      D
A 0.9987969 0.0002823557 0.0009207256
I 0.0000000 0.9320343984 0.0679656016
D 0.0000000 0.0000000000 1.0000000000

Markovchain 3
vek 42 Prechodova matica
A 3 - dimensional discrete Markov Chain defined by the following states:
A, I, D
The transition matrix (by rows) is defined as follows:
      A      I      D
A 0.9987003 0.0003147021 0.0009849912
I 0.0000000 0.9242689188 0.0757310812
D 0.0000000 0.0000000000 1.0000000000

```

Tento objekt je dôležitý na realizáciu simulácií. Teraz nasimulujeme 100 000 životných trajektórií pre osobu vo veku 40 rokov. Táto simulácia je vytvorená pomocou (6.4).

```

prechody<-rmarkovchain(n = 100000, object = kompletnaPrechodM(a),
                        what = "matrix", t0 = "A")

```

Počiatočný stav t0 je nastavený tak, aby 40 ročná osoba bola zdravá. To je predpoklad k tomu aby sa mohla dať poistiť. Výstupom je 100 000 simulácií pre osobu 40 rokov.

```

      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12]
[1,] "A"  "A"  "A"  "A"  "A"  "A"  "A"  "A"  "A"  "A"  "A"  "A" ...
[2,] "A"  "A"  "A"  "A"  "A"  "A"  "A"  "A"  "A"  "A"  "A"  "A" ...
[3,] "A"  "A"  "A"  "A"  "A"  "A"  "A"  "A"  "A"  "A"  "A"  "A" ...
[4,] "A"  "A"  "A"  "A"  "A"  "A"  "A"  "A"  "A"  "A"  "A"  "A" ...
[5,] "A"  "A"  "A"  "A"  "A"  "A"  "A"  "A"  "A"  "A"  "A"  "A" ...

      [,68] [,69] [,70] [,71] [,72] [,73] [,74] [,75] [,76] [,77] [,78]
[1,] "D"  "D"  "D"  "D"  "D"  "D"  "D"  "D"  "D"  "D"  "D"
[2,] "D"  "D"  "D"  "D"  "D"  "D"  "D"  "D"  "D"  "D"  "D"
[3,] "D"  "D"  "D"  "D"  "D"  "D"  "D"  "D"  "D"  "D"  "D"
[4,] "D"  "D"  "D"  "D"  "D"  "D"  "D"  "D"  "D"  "D"  "D"
[5,] "D"  "D"  "D"  "D"  "D"  "D"  "D"  "D"  "D"  "D"  "D"

      [,79] [,80]
[1,] "D"  "D"
[2,] "D"  "D"
[3,] "D"  "D"
[4,] "D"  "D"
[5,] "D"  "D"

[ reached getOption("max.print") -- omitted 99988 rows ]

```

Ako vidíme z dát máme tam 80 hodnôt. Čo vyplýva z toho, že naša osoba je vo veku 40 a naše tabuľky sú do veku 120. Z toho vyplýva, že musí byť 80 hodnôt v jednom riadku.

V nasledujúcej časti chceme odhadnúť priemerný čas, za aký poistenec vo veku 40 rokov ochorie. Na to použijeme nasimulované dáta. Vytvoríme si nulovú maticu, ktorá bude mať rovnaké rozmery ako matica nasimulovaných dát. Do nulovej matice sa priradí jednotka na to isté miesto, na ktorom bola v nasimulovanej matici (A). Následne sa jednotky v riadku sčítajú a potom sa všetky prvky spriemerujú. V softvéri R použijeme na to tieto príkazy.

```
tem<-matrix(0,nrow=nrow(prechody),ncol = ncol(prechody))
tem[prechody=="A"]<-1
tem
cas_v_A<-mean(rowSums((tem)))
cas_v_A
```

Výstupom je matica jednotiek. Z nej vieme následne vypočítať priemerný čas za ktorý poistenec ochorie. Tento čas je v rokoch.

```
      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12]
[1,]    1    1    1    1    1    1    1    1    1    1    1    1 ...
[2,]    1    1    1    1    1    1    1    1    1    1    1    1 ...
[3,]    1    1    1    1    1    1    1    1    1    1    1    1 ...
[4,]    1    1    1    1    1    1    1    1    1    1    1    1 ...
[5,]    1    1    1    1    1    1    1    1    1    1    1    1 ...
      [,68] [,69] [,70] [,71] [,72] [,73] [,74] [,75] [,76] [,77] [,78]
[1,]     0     0     0     0     0     0     0     0     0     0     0
[2,]     0     0     0     0     0     0     0     0     0     0     0
[3,]     0     0     0     0     0     0     0     0     0     0     0
[4,]     0     0     0     0     0     0     0     0     0     0     0
[5,]     0     0     0     0     0     0     0     0     0     0     0
      [,79] [,80]
[1,]     0     0
[2,]     0     0
[3,]     0     0
[4,]     0     0
[5,]     0     0

[ reached getOption("max.print") -- omitted 99988 rows ]
[1] 39.85612
```

Takže čas za ktorý poistenec ochorie je priemerne 39,85 rokov. Analogicky vieme vypočítať aj čas ako dlho bude osoba v stave chorý. Tento čas je 1,44. Tieto čísla budeme potrebovať pri nasledujúcej analýze.

Teraz budeme predpokladať, že poistenec sa chce dať poistiť tak, že ak sa dostane do stavu nesebestačnosti, chce aby sa mu následne päť rokov vyplácalo 500 € mesačne pri úrokovej miere 1,2 %. V tomto prípade neberieme do úvahy náklady poisťovne. Hodnotu, ktorú musí mať poistenec nasporenú pred tým ako ochorie vieme vypočítať pomocou základného finančného vzorca. A_n (Cipra 1995)

$$A_n = R \frac{1 - \left(1 + \frac{i^{(m)}}{m}\right)^{-mn}}{\left(1 + \frac{i^{(m)}}{m}\right)^{\frac{m}{p}} - 1}$$

A_n je suma, ktorú musí mať poistenec nasporenú keď ochorie

R je renta, ktorú chce poistenec dostávať ak ochorie

i je úroková miera

m je počet konverzií za rok

p je počet splátok ročne

n je počet rokov

$$A_5 = 500 \frac{1 - \left(1 + \frac{0,012^{(1)}}{1}\right)^{-5}}{\left(1 + \frac{0,012^{(1)}}{1}\right)^{\frac{1}{12}} - 1}$$

$$A_5 = 29\,108,41$$

Sumu, ktorú musí mať poistenec nasporenú je 29 108,41 €. Na takéto poistenie si chce sporiť 20 rokov. Sumu, ktorú musí zaplatiť dnes vieme vypočítať pomocou štandardného finančného vzorca.

$$K_0 = K_n \left(1 + \frac{i^{(m)}}{m}\right)^{-mn}$$

K_0 je suma, ktorú musí teraz zaplatiť poistenec

K_n je suma, ktorú si chce nasporiť

$$K_0 = 29\,108,41 \left(1 + \frac{0,012^{(1)}}{1}\right)^{-20}$$

$$K_0 = 22\,930,22$$

Poistenec musí teraz zaplatiť 22 930,22 €. Ak by poisťovňa kalkulovala s časmi, ktoré si vie odhadnúť pomocou simulácií. Podľa simulácií nám vyšiel priemerný čas, za ktorý bude poistenec chorý 1,44. Suma, ktorú by si mal poistenec nasporiť pred tým ako ochorie, by bola 8 561, 96 €. Priemerný čas, ktorý bude poistenec zdravý je 39,85. Podľa toho by stačilo, aby poistenec pri podpise zaplatil 5 322,67 €. Ako vidíme poisťovni tak vzniká veľký priestor na zisk.

Pomocou simulácií sa dá vypočítať poistné aj iným spôsobom. Predpokladajme, že poistenec chce, aby mu poisťovňa vyplatila pri chorobe 6500 € ročne pri úrokovej miere 1,2 %. Použijeme maticu jednotiek, ktorú vynásobíme sumou 6500 a stĺpcovým vektorom úrokových mier a následne všetky prvky spriemerujeme.

```
head(6500*temp%*( matrix((1+0.012)^~seq(from=0, to=ncol(temp)-1))))
mean(6500*temp%*( matrix((1+0.012)^~seq(from=0, to=ncol(temp)-1))))
```

Výstupom je stĺpcový vektor, ktorý následne spriemerujeme a dostaneme koľko musí mať poistenec vo veku 40 rokov nasporené, aby mu mohla poisťovňa vyplácať 6500 € ročne.

```
      [,1]
[1,]    0.00
[2,] 10740.63
[3,]    0.00
[4,]    0.00
[5,]    0.00
[6,] 10118.75
[ reached getOption("max.print") -- omitted 99000 rows ]
[1] 5869.701
```

Poistenec by mal mať nasporených minimálne 5 869,70 €.

4.3. Diskusia

Markovové reťazce majú široké využitie v aktuárskej praxi. V našej práci sme sa zamerali na aplikáciu Markovových reťazcov na dve základné sektory.

Prvý sektor bol z oblasti neživotného poistenia, konkrétnejšie išlo o povinné zmluvné poistenie zodpovednosti za škodu spôsobenú vedením motorového vozidla. Software R, v ktorom sme pracovali nám poskytol balíček `markovchain`. Tento balíček obsahuje všetky potrebné metódy, ktoré sú potrebné pri riešení problematiky PZP a bližšie pre bonus malus systém. Ak by sa poisťovňa riadila týmto modelom, musela by základné poistné zvoliť tak aby aj po tom ako priemerne od poistencov prijme 54,32 % zo základného

poistného zostala solventná. V našom modeli nie sú zahrnuté náklady poisťovne, ktoré tvoria vysokú časť výdavkov poisťovne. V skutočnosti však poisťovne nepoužívajú takýto jednoduchý systém. Do výšky poistného vstupuje aj vek poistenca, miesto bydliska a napokon aj škodové udalosti. V našom prípade sme brali do úvahy iba škodovať poistenca. Naša frekvencia škodovosti predpokladala to, že priemerný vodič spôsobí jednu poistnú udalosť za 10 rokov. No poisťovne majú k dispozícii dáta za dlhé obdobia. V tomto prípade by odhady matice prechodu boli realistickejšie. Potom by aj vektor absolútnych pravdepodobností mal lepšiu výpovednú hodnotu. V skutočnosti sa systém nikdy nedostane do tohto stavu. Je to však najlepšia aproximácia tohto stavu v dlhodobom horizonte.

Druhý sektor bol z oblasti životného a nemocenského poistenia. V tomto prípade boli naše dáta z talianskej populácie a to iba pre mužskú populáciu. Dáta sa zaoberajú pravdepodobnosťami prechodu do stavu nesebestačnosti. Čo je stav spôsobený chorobou alebo úrazom. Človek je nesebestačný, ak nedokáže vykonávať niektorú z činností ako je neschopnosť prijímania stravy, umývania sa, výkonu fyziologickej potreby, obliekania sa, strata mobility a schopnosti pohybu. Ak by poisťovňa sledovala takéto ochorenia a ich prechody, vedela by sa pripraviť na vyplácanie dávok. V opačnom prípade by mohla pozorovať čas zotrvania v jednotlivých stavoch a tak svojich poistencov vyzývať k prevencii. Tým by mohla predísť k prechodu do takéhoto stavu a znížiť svoje výdavky na vyplácanie poistného plnenia alebo renty. V súčasnosti už niektoré poisťovne aplikujú túto možnosť na svojich poistencov. Vyžívajú ich hlavne k prevencií pred rakovinou hrubého čreva a srdcovým ochoreniam. Robia to hlavne cestou všeobecných lekárov,, ktorí pacientov k tomu nabádajú a informujú ich.

Software R je veľmi užitočný nástroj na riešenie štatistických problémov. Je v ňom mnoho užitočných nástrojov a balíkov, ktoré sa dajú aplikovať na problematiku Markovových reťazcov. Ako sme mohli vidieť v ukážkach, tak pri inštalácii balíka sa nám ponúкло široké množstvo jednoduchých príkazov, ktoré nám riešili celkom zložité situácie. Za takýto balík môžeme poďakovať poprednému štatistikovi Giorgio Alfredo Spedicato. Zaoberá sa implementáciu štatistických metód do software R.

Záver

Cieľom tejto práce bolo zoznámiť sa s teóriou Markovových reťazcov a ukázať ich použitie v oblasti aktuárstva. V práci sa nachádza aj teoretický základ, ktorý je potrebný pri riešení danej problematiky. Boli popísané a definované všetky typy Markovových reťazcov. V práci sme sa však venovali predovšetkým homogénemu Markovovmu reťazcu s diskretným časom. Spojitému času sme sa venovali iba v teoretickej rovine. Boli predstavené základné operácie Markovových reťazcov a ich aplikácia na poistenie.

Následne bol opísaný vybraný balíček `markovchain` v software R. Boli predstavené všetky metódy, ktoré sme použili v praktickej časti a ich demonštrácia na trojstavovom modeli. Tento model bol zvolený kvôli jeho jednoduchosti, aby sa na ňom dali demonštrovať všetky metódy.

Aplikácia bola uskutočnená v oboch oblastiach poistenia aj v životnej aj neživotnej oblasti poistenia. V oblasti neživotného poistenia bola naša aplikácia na bonus malus systém. V tejto časti sme ukázali vytvorenie matice pravdepodobnosti prechodu pomocou škodovej frekvencie, ktorá vychádzala z predpokladu, že poistenec spôsobí jednu poistnú udalosť za 10 rokov. Matica pravdepodobnosti prechodu sa riadila podľa pravidiel prechodu medzi stavmi, ktoré sme definovali ako prvé. Následne sme ukázali výpočet vektora absolútnych pravdepodobností a jeho využitie pre poisťovne.

V oblasti životného poistenia bola naša aplikácia na dátach o dlhodobo chorých. Tieto informácie boli o mužskej populácii v Taliansku za rok 2016. Dáta boli zostavené od 20. do 120. veku osoby v populácii. Následne sme nasimulovali možné životné trajektórie 40 ročného poistenca, ktorého počiatočný stav bol zdravý. Následne sme vypočítali poistné na základne požiadaviek poistenca. Následne sme ich porovnali s poistným, ktoré nám vyšlo zo simulácií. Poistné by sa mu vyplatilo, ak by sa dostal do stavu dlhodobo chorý. V tomto prípade išlo o nehomogénny Markovov reťazec. Do tohoto reťazca nevstupoval iba čas ale aj vek poistenca. A to znamená, že matica pravdepodobnosti prechodu sa s časom mení.

Zoznam použitej literatúry

Knižné zdroje:

CIPRA, Tomáš: *Průvodce finanční a pojistnou matematikou*, 1995

DICKSON, D., HARDY, M., WATERS, H.: *Actuarial Mathematics for Life Contingent Risks*, 2009

ETHIER, S. E. and KURTZ, T.G.: *Markov Processes Characterization and Convergence*. Wiley, Series in Probability and Mathematical Statistics, New York, 1986.

FECENKO, Jozef: *Neživotné poistenie*, Bratislava, 2012

GRINSTEAD, C. M. and SNELL, J. L.: *Introduction to Probability* [online]. 1997. Dostupný NA WWW http://www.dartmouth.edu/~chance/teaching_aids/books_articles/probability_book/book.html

HABERMAN S., PITACCO E.: *Actuarial models for disability insurance*, 1CHUNG, K. L.: *Markov Chains with Stationary Transition Probabilities*. Springer Verlag, New York, 1967.

HAVRDA, J.: *Náhodné procesy*, Praha, 1981

JANKOVÁ, K., BRUNOVSKÝ, P., KILIANOVÁ, S., BOKES, P.: *Markovove reťazce a ich aplikácie*, Slovenská republika, 2014.

KNILL, O.: *Probability and Stochastic Processes with Applications* [online]. 2011. Dostupný na WWW <http://www.math.harvard.edu/~knill/books/KnillProbability.pdf>

KŘIVÝ, I.: *Náhodné procesy*. Ostravská univerzita, Ostrava, 2005.

PRÁŠKOVÁ, Z., LACHOUT, P.: *Základy náhodných procesů I.*, Karolinum, Praha, 2005.

SPEDICATO, Giorgio Alfredo, KANG, Tae Seung, YALAMANCHI, Sai Bhargav, YADAV, Deepak. *The markovchain Package: A Package for Easily Handling Discrete Markov Chains in R*, 2018.

SPEDICATO, Giorgio Alfredo: *Discrete Time Markov Chains with R*, 2017

SEKERKOVÁ, Viera a BILÍKOVÁ, Márie, *Poistná Matematika Bratislava*, 2005.

ŠTEPÁN, J.: *Teorie pravdě podobnosti*, Praha, 1987.

ŠKROVÁNKOVÁ Lea, *Zdravotné a nemocenské poistenie*, Bratislava, 2013.

Internetové zdroje:

- [14] https://fhi.euba.sk/www_write/files/katedry/ks/veda-vyskum/2015_6_3_5_seminar_doktorandov_FHI_KME_2015_zbornik.pdf
(24.4.2019)
- [15] <https://www.alphamedical.sk/casopis-invitro/tuberkuloza-stale-znamy-nepriatel>
(24.4.2019)
- [16] <http://rusnak.truni.sk/prednasky/Statistika%20zdravotnicka/Standardizacia/standardizacia.pdf> (24.4.2019)

Prílohy

Príloha č.1:

```
BMretazec12 <- function(lambda)
{
  bmM12 <- matrix(0, 12, 12)

  bmM12[1, 1] <- dpois(x = 0, lambda)
  bmM12[1, 3] <- dpois(x = 1, lambda)
  bmM12[1, 5] <- dpois(x = 2, lambda)
  bmM12[1, 7] <- dpois(x = 3, lambda)
  bmM12[1, 9] <- dpois(x = 4, lambda)
  bmM12[1, 11] <- dpois(x = 5, lambda)
  bmM12[1, 12] <- 1 - ppois(q = 5, lambda)

  bmM12[2, 1] <- dpois(x = 0, lambda)
  bmM12[2, 4] <- dpois(x = 1, lambda)
  bmM12[2, 6] <- dpois(x = 2, lambda)
  bmM12[2, 8] <- dpois(x = 3, lambda)
  bmM12[2, 10] <- dpois(x = 4, lambda)
  bmM12[2, 12] <- 1 - ppois(q = 4, lambda)

  bmM12[3, 2] <- dpois(x = 0, lambda)
  bmM12[3, 5] <- dpois(x = 1, lambda)
  bmM12[3, 7] <- dpois(x = 2, lambda)
  bmM12[3, 9] <- dpois(x = 3, lambda)
  bmM12[3, 11] <- dpois(x = 4, lambda)
  bmM12[3, 12] <- 1 - ppois(q = 4, lambda)

  bmM12[4, 3] <- dpois(x = 0, lambda)
  bmM12[4, 6] <- dpois(x = 1, lambda)
  bmM12[4, 8] <- dpois(x = 2, lambda)
  bmM12[4, 10] <- dpois(x = 3, lambda)
  bmM12[4, 12] <- 1 - ppois(q = 3, lambda)

  bmM12[5, 4] <- dpois(x = 0, lambda)
  bmM12[5, 7] <- dpois(x = 1, lambda)
  bmM12[5, 9] <- dpois(x = 2, lambda)
  bmM12[5, 11] <- dpois(x = 3, lambda)
  bmM12[5, 12] <- 1 - ppois(q = 3, lambda)

  bmM12[6, 5] <- dpois(x = 0, lambda)
  bmM12[6, 8] <- dpois(x = 1, lambda)
  bmM12[6, 10] <- dpois(x = 2, lambda)
  bmM12[6, 12] <- 1 - ppois(q = 2, lambda)

  bmM12[7, 6] <- dpois(x = 0, lambda)
  bmM12[7, 9] <- dpois(x = 1, lambda)
  bmM12[7, 11] <- dpois(x = 2, lambda)
  bmM12[7, 12] <- 1 - ppois(q = 2, lambda)
```

```

bmM12[8, 7] <- dpois(x = 0, lambda)
bmM12[8, 10] <- dpois(x = 1, lambda)
bmM12[8, 12] <- 1 - ppois(q = 1, lambda)

bmM12[9, 8] <- dpois(x = 0, lambda)
bmM12[9, 11] <- dpois(x = 1, lambda)
bmM12[9, 12] <- 1 - ppois(q = 1, lambda)

bmM12[10, 9] <- dpois(x = 0, lambda)
bmM12[10, 12] <- 1 - ppois(q = 0, lambda)

bmM12[11, 10] <- dpois(x = 0, lambda)
bmM12[11, 12] <- 1 - ppois(q = 0, lambda)

bmM12[12, 11] <- dpois(x = 0, lambda)
bmM12[12, 12] <- 1 - ppois(q = 0, lambda)

stavyBM12 <- as.character(1:12)
BM12 <- new("markovchain", transitionMatrix = bmM12,
            states = stavyBM12, name = "BM Matica")
return(BM12)
}

BMM12 <- BMretazec12(0.1)
BMM12

```

Príloha č.2:

```

plot(BMM12)

stochastic_matrix_to_plot <- as(BMM12,"matrix")
plotmat(stochastic_matrix_to_plot,lwd = 1, box.lwd = 1,
        cex.txt = 0.5,box.size = 0.05,box.type = "circle",
        box.prop = 0.5,box.col = "light blue",arr.length=.1,
        arr.width=.1,self.cex = .6,main = "Bonus - Malus systém")

```